PolyPublie
Polytechnique Montréal

POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

Titre: Title:	Traitement paramétrique des signaux audio dans le contexte des prothèses auditives
Auteur: Author:	Abdelaziz Trabelsi
Date:	2008
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Trabelsi, A. (2008). Traitement paramétrique des signaux audio dans le contexte des prothèses auditives [Thèse de doctorat, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/8201/

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/8201/
Directeurs de recherche: Advisors:	
Programme: Program:	Non spécifié

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal

https://publications.polymtl.ca

UNIVERSITÉ DE MONTRÉAL

TRAITEMENT PARAMÉTRIQUE DES SIGNAUX AUDIO DANS LE CONTEXTE DES PROTHÈSES AUDITIVES

ABDELAZIZ TRABELSI

DÉPARTEMENT DE GÉNIE INFORMATIQUE ET GÉNIE LOGICIEL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION

DU DIPLÔME DE PHILOSOPHIAE DOCTOR

(GÉNIE INFORMATIQUE)

DÉCEMBRE 2008

© Abdelaziz Trabelsi, 2008.

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-48892-8
Our file Notre référence
ISBN: 978-0-494-48892-8

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

TRAITEMENT PARAMÉTRIQUE DES SIGNAUX AUDIO DANS LE CONTEXTE DE PROTHÈSES AUDITIVES

Présentée par : TRABELSI Abdelaziz

en vue de l'obtention du diplôme de : Philosophiae Doctor

a été dûment acceptée par le jury d'examen constitué de :

Mme. CHERIET Farida, Ph.D, présidente

M. BOYER François-Raymond, Ph.D, membre et directeur de recherche

M. SAVARIA Yvon, Ph.D, membre et codirecteur de recherche

M. SAUCIER Antoine, Ph.D, membre

M. BÉGIN Guy, Ph.D, membre

DÉDICACE

À toute ma famille

Je vous aime.

REMERCIEMENTS

Cette page est dédiée à tous ceux que je voudrais remercier pour m'avoir soutenu, encouragé ou accompagné durant mes années de thèse.

Je tiens à remercier chaleureusement mon directeur de thèse M. François-Raymond BOYER, professeur au département de génie informatique et génie logiciel de l'École Polytechnique de Montréal, et mon codirecteur M. Yvon SAVARIA, professeur au département de génie électrique de l'École Polytechnique de Montréal, pour avoir aimablement accepté de diriger cette thèse. Leurs conseils précieux, leur soutien permanent, leur patience et leur implication m'ont permis de progresser et de consolider la rigueur de ce travail de thèse. Qu'ils trouvent ici l'expression de ma vive gratitude et mon profond respect.

Je remercie Mme Farida CHERIET, professeure au département de génie informatique et génie logiciel de l'École Polytechnique de Montréal, pour avoir bien voulu présider le jury de cette thèse. J'aimerais aussi remercier M. Antoine SAUCIER, professeur au département de mathématiques et de génie industriel de l'École Polytechnique de Montréal, et M. Guy BÉGIN, professeur au département d'informatique de l'Université du Québec à Montréal, qui ont accepté de faire partie de ce jury. Les corrections et les remarques qu'ils ont apportées à ce document ont été particulièrement enrichissantes.

Je suis particulièrement reconnaissant à M. Mounir BOUKADOUM, professeur au département d'informatique de l'Université du Québec à Montréal, pour son aide inestimable et l'immense sympathie qu'il a toujours manifesté à mon égard. Ses qualités scientifiques et humaines, son encouragement et ses nombreux et précieux conseils ont largement contribué à l'aboutissement de cette thèse. Je le remercie également pour m'avoir confié des enseignements. Qu'il trouve avec ceci un modeste geste de reconnaissance et de remerciement.

J'adresse mes remerciements à Mme Ghyslaine ETHIER-CARRIER, secrétaire au département de génie électrique de l'École Polytechnique de Montréal, pour ses qualités humaines et pour m'avoir aidé à organiser ma participation à la conférence ICECS2007.

Je souhaite également remercier tous ceux qui ont participé aux nombreux tests d'écoutes, en particulier M. Bruno GIRODIAS, étudiant au doctorat au département de génie informatique et génie logiciel de l'École Polytechnique de Montréal. Merci Bruno!

Enfin, je suis spécialement reconnaissant envers mes parents, ma femme Teodora, ma fille Nadia, mon fils Karim, mes frères et mes amis, qui ont toujours été présents. Merci pour votre soutien affectif en toute circonstance.

RÉSUMÉ

Les aides auditives numériques pour malentendants constituent un champ d'application particulièrement important du traitement du signal. Les dernières générations d'aides numériques sont équipées d'un réseau local sans fil permettant un véritable traitement binauriculaire de l'information échangée non seulement pour améliorer son intelligibilité et son confort d'écoute, mais aussi pour renforcer les indices acoustiques de localisation de la source émettrice. Ce progrès technologique a rendu possible l'utilisation des techniques de traitement paramétrique classique qui sont largement utilisées pour la compression de parole en téléphonie.

Le problème de la sensibilité au bruit de fond des traitements paramétriques est bien connu. Deux approches de traitement paramétrique en présence de bruit sont souvent utilisées. La première approche consiste à effectuer un traitement pour la réduction de bruit, généralement dans le domaine fréquentiel, comme un prétraitement pour la modélisation paramétrique d'un signal audio. Quant à la seconde approche, elle consiste à appliquer un traitement pour la réduction de bruit directement dans le domaine de corrélation afin d'évaluer les paramètres spectraux qui modélisent convenablement la structure formantique du signal. Nous proposons des solutions algorithmiques innovantes et pertinentes pour déterminer laquelle des deux approches permet d'obtenir la meilleure fidélité de l'enveloppe spectrale, d'un signal audio, à sa structure formantique.

Dans le cadre de la première approche, nous étudions une méthode de filtrage adaptatif pour la réduction de bruit, proposée auparavant dans la littérature, et nous identifions quelques problèmes inhérents associés à son utilisation. Pour pallier ces problèmes, nous proposons de modifier un estimateur existant de densité spectrale de puissance du bruit et le combiner à cette méthode de filtrage au moyen d'une procédure de décision quantifiée. Nous montrons que cette combinaison est avantageuse dans la mesure où elle permet un bon compromis entre la réduction de bruit et la distorsion produite sur le signal d'intérêt.

Concernant la seconde approche, nous proposons deux versions d'une méthode itérative et originale de traitement paramétrique qui repose sur un modèle linéaire autorégressif, une condition d'arrêt prédéfinie et l'algorithme de MS (pour « Minimum Statistics »). Nous montrons que cette méthode, qui opère dans le domaine de corrélation, permet d'obtenir des paramètres spectraux compensés et stables. Nous présentons les limites associées à l'utilisation de l'algorithme de MS dans le domaine de corrélation. Ensuite, une méthode permettant d'améliorer la précision de l'estimateur de la variance du bruit dans le cas d'un signal audio contaminé par un bruit blanc additif a été développée. En comparaison avec la méthode qui repose sur l'algorithme de MS, nous montrons empiriquement que cette nouvelle méthode permet d'obtenir des paramètres spectraux plus robustes et qui modélisent convenablement la structure formantique du signal.

ABSTRACT

The digital hearing aids for the hearing-impaired constitute a particularly important application field of signal processing. The last generations of hearing aids are equipped with a wireless local area network which allows a sophisticated binaural processing of the exchanged information not only to improve its intelligibility and its quality perception but also to emphasize the associated source localization acoustic cues. This technological progress has made possible the use of conventional parametric processing techniques which are widely used for speech compression in telephony.

The sensitivity problem of the conventional parametric processing techniques to background noise is well-known. Two approaches of parametric processing in the presence of noise are often used. The first approach consists of performing the reduction of noise, generally in the frequency domain, as a preprocessing for the parametric modeling of the audio signal. The second approach allows applying a noise reduction processing directly in the correlation domain. This processing allows estimating the spectral parameters which model suitably the formantic structure of the signal. We propose innovative and relevant algorithmic solutions to determine which of both approaches allows obtaining the best fidelity of the spectral envelope of an audio signal to its formantic structure.

Within the context of the first approach, we study an adaptive filtering method for noise reduction, proposed previously in the literature, and we identify some inherent problems associated with its use. To fix these problems, we suggest modifying an

existing noise power spectrum estimator and combining it with this filtering method by means of a soft-decision scheme. We show that this combination is advantageous in that it provides a good tradeoff between the amount of noise reduction and the speech distortion.

Regarding the second approach, we propose two versions of an iterative method for parametric processing based on an autoregressive linear model, a predefined decision criterion and the Minimum Statistics (MS) algorithm. We show that this method, which operates in the correlation domain, allows obtaining compensated and stable spectral parameters. We present the limits associated with the use of the MS algorithm in the correlation domain. Then, a method that allows improving the precision of the noise variance estimator in the case of an audio signal contaminated by an additive white noise was developed. In comparison with the method based on the MS algorithm, we show empirically that the new proposed method gives rise to more robust spectral parameters which suitably model the formantic structure of the signal.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

LISTE DES SYMBOLES ET ABRÉVIATIONS

Pour des raisons de clarté, la signification d'un symbole ou d'une abréviation n'est souvent rappelée qu'à sa première apparition dans le texte d'un chapitre.

AR Autoregressive.

ARMA Autoregressive Moving-Average.

ASA American Standard Association.

CSS Cross-Spectral Subtraction.

DISP Densité Interspectrale de Puissance.

DNF Diffuse Noise Field.

DSP Densité Spectrale de Puissance (terme en Français de PSD pour « Power Spectral Density »).

HINT Hearing in Noise Test (database)

HOYWE High-Order Yule-Walker Equations.

LOYWE Low-Order Yule-Walker Equations.

LPC Linear Prediction Coding.

LS Least Squares.

MA Moving-Average.

MEM Maximum Entropy Method.

MLSE Maximum Likelihood Spectral Estimation.

MS Minimum Statistics.

ODNE Overdetermined Normal Equations.

PSD Power Spectral Density.

RPE Recursive Prediction Error.

RSB Rapport Signal sur Bruit (terme en Français de SNR pour « Signal to Noise Ratio »).

SIS Soustraction Interspectrale (terme en Français de CSS pour « Cross-Spectral Subtraction »).

SNR Signal to Noise Ratio.

SVD Singular Value Decomposition.

TSVD Truncated Singular Value Decomposition.

VAD Voice Activity Detector.

ω Fréquence en radians par seconde.

N Largeur de la fenêtre d'analyse.

D Largeur de la fenêtre de recherche du minimum.

p Ordre d'un modèle autorégressif.

LISTE DES ANNEXES

CHAPITRE 1
INTRODUCTION

1.1 Problématique

La compréhension de la parole en situation bruyante est souvent difficile pour un individu à l'audition normale. Cette situation est encore plus difficile pour un individu malentendant (atteint d'une déficience auditive ou surdité). La malentendance a pour conséquence de porter atteinte à la qualité de vie d'un individu malentendant en affectant non seulement son bien-être émotionnel et psychologique, mais aussi social et physique. Le port d'aides auditives est bien souvent essentiel pour les individus atteints de certaines formes de surdités, notamment sévères ou profondes (classées selon le degré de perte auditive).

Les aides auditives numériques pour malentendants constituent un champ d'application particulièrement important du traitement du signal. L'intégration d'algorithmes de calculs numériques sophistiqués dans ces aides auditives est une orientation technologique assez récente qui a été rendue possible grâce à la miniaturisation des processeurs de traitement du signal. Les dernières générations d'aides auditives numériques sont équipées d'un réseau local permettant un véritable traitement binauriculaire du signal (comme l'échange d'information audio large bande entre deux aides auditives par exemple) non seulement pour améliorer son intelligibilité et son confort d'écoute, mais aussi pour rehausser les indices acoustiques de localisation de la source émettrice (indices binauriculaires). Ce progrès technologique a rendu

possible l'utilisation des techniques de traitement paramétrique classique qui sont largement utilisées pour la compression de parole en téléphonie. Ce sont des procédures de compression habituellement appelées « techniques de codage ».

Le problème de la sensibilité au bruit de fond des traitements paramétriques est bien connu. Lors de l'analyse, de faibles variations entre deux trames de données consécutives peuvent entraîner une déviation importante des paramètres spectraux par rapport à ceux extraits du signal original. Ces variations sont habituellement générées soit par un bruit ambiant, soit par l'erreur de quantification. À la synthèse, cette déviation entraîne de fortes variations dans le spectre restitué par le filtre de reconstruction. Ce phénomène d'instabilité conduit souvent à une dégradation globale de la qualité de perception du signal audio reconstruit. Pour pallier ce problème, l'information échangée subit, outre le traitement paramétrique classique, un traitement spécifique pour la réduction du bruit de fond. L'ordre dans lequel ces deux traitements sont appliqués est important. Dans le cadre de ce travail de thèse, l'avantage d'effectuer un traitement pour la réduction de bruit suivi d'un traitement paramétrique classique, ou vice versa, est examiné en termes de fidélité de l'enveloppe spectrale, de l'information échangée, à sa structure formantique.

1.2 Cadre et objectifs de la thèse

Ce projet de thèse porte sur le traitement numérique des signaux dans le contexte de la compensation du handicap chez les malentendants par port d'aides auditives. Nous considérons le cas des aides auditives numériques dotées d'un réseau local de

communication. L'objectif principal de ce travail de thèse est de proposer des solutions algorithmiques innovantes et pertinentes pour améliorer la qualité de l'information audio large bande qui pourra être échangée entre deux aides auditives.

Le modèle paramétrique d'un signal audio comporte typiquement quatre paramètres : le voisement, la fréquence fondamentale ou « pitch », l'évolution temporelle de l'énergie du signal, et l'enveloppe spectrale du signal. Dans ce travail de thèse, nous nous intéressons exclusivement à l'ajustement de l'enveloppe spectrale d'un signal audio dégradé par la présence de bruit. L'enveloppe spectrale d'un signal peut être obtenue par une analyse de Fourier à court terme synchrone avec la fréquence fondamentale ou par modélisation linéaire autorégressive au moyen d'un filtre de prédiction linéaire. Notre choix s'est porté sur la modélisation linéaire prédictive dite souvent analyse LPC pour « Linear Predictive Coding ».

Le problème de la sensibilité au bruit de fond des traitements paramétriques est bien connu. Il faudrait alors prévoir un traitement spécifique pour la réduction de bruit. Deux techniques de traitement paramétrique en présence de bruit sont souvent utilisées :

1. La première technique consiste à effectuer un traitement pour la réduction de bruit, généralement dans le domaine fréquentiel, comme un prétraitement pour la modélisation paramétrique du signal audio. Bien que cette approche soit efficace pour contourner les problèmes d'instabilité souvent rencontrés par les techniques de modélisation linéaire prédictive, elle est susceptible de générer des tons musicaux qui peuvent dégrader la performance de l'analyse spectrale par LPC.

2. La seconde technique consiste plutôt à appliquer un traitement pour la réduction de bruit directement dans le domaine de corrélation afin d'évaluer les paramètres spectraux qui modélisent convenablement la structure formantique du signal. Bien que cette approche soit particulièrement robuste à l'ajustement de l'enveloppe spectrale d'un signal audio dégradé par la présence de bruit, elle ne permet en effet qu'une réduction partielle du bruit décidée par la contrainte de stabilité des paramètres spectraux à évaluer. Il faudrait évidemment déterminer laquelle des deux techniques permet d'obtenir la meilleure fidélité de l'enveloppe spectrale, d'un signal audio, à sa structure formantique.

1.3 Contributions et originalité

Les travaux de recherche, en traitement des signaux audio, couverts par cette thèse ont fait l'objet de quatre contributions particulièrement pertinentes.

La première contribution correspond au développement, dans le domaine fréquentiel, d'une nouvelle approche à deux microphones basée sur la combinaison d'une technique de filtrage adaptatif et d'un estimateur de densité spectrale de puissance (DSP) du bruit. L'estimateur de la DSP du bruit proposé repose sur une version rapide de la technique de MS (pour « Minimum Statistics »), que nous avons implémentée, et une procédure de décision quantifiée. Cet estimateur possède deux fonctions essentielles: rendre inaudible le bruit résiduel généré par le filtrage adaptatif et réduire le bruit cohérent (particulièrement en basse fréquence) nécessairement présent dans un champ lointain (« Diffuse Noise Field »). Cette combinaison permet ainsi un meilleur compromis entre

la réduction de bruit et la distorsion produite sur le signal d'intérêt. Pour évaluer la méthode proposée en termes de performance, nous avons réalisé de nombreux tests objectifs et subjectifs sur plusieurs signaux audio contaminés par différents types de bruit pour différents RSB. Les résultats de ces tests ont démontré la supériorité de notre méthode par rapport à d'autres méthodes concurrentes, notamment dans des situations de bruit fortement non stationnaire (bruit impulsif et bruit d'ambiance de type « cocktail-party »). De plus, la complexité réduite de la méthode permet d'envisager une implémentation en temps réel du système combiné. Cette approche a fait l'objet d'un rapport technique [47], un article de conférence [48] et un article de revue [49].

La seconde contribution porte sur l'élaboration, dans le domaine de corrélation, de deux versions d'une approche originale qui permet de limiter la distorsion de l'enveloppe spectrale d'un signal audio large bande corrompu par un bruit de fond additif. Il s'agit d'une méthode itérative qui repose sur une condition d'arrêt prédéfinie (la matrice de corrélation compensée doit être définie positive) et l'algorithme de MS. Des paramètres spectraux compensés et stables peuvent ainsi être obtenus aussi longtemps que les coefficients de réflexion estimés sont strictement inférieurs à l'unité en amplitude. La méthode est particulièrement appropriée aux bruits de fond dont les effets s'étendent sur l'ensemble des retards de la fonction de corrélation. Ces deux versions ont été évaluées expérimentalement et comparées l'une à l'autre. La première version de cette approche ne permet en effet qu'une faible réduction de bruit due à la contrainte d'obtenir, à chaque itération, une matrice de corrélation définie positive pour l'ensemble des retards de corrélation compensés. Quant à la seconde version, elle fournit

une meilleure réduction de bruit en permettant de compenser individuellement les retards de corrélation des effets d'un bruit de fond. Cette approche a fait l'objet de deux articles de conférence [50,51].

Élaboré dans le domaine de corrélation, un nouvel estimateur de la variance du bruit a été développé et constitue notre troisième contribution. Cet estimateur repose sur la méthode ODNE de Cadzow [1], la décomposition en valeurs singulières (SVD pour « Singular Value Decomposition ») tronquée de la matrice de corrélation (au sens des moindres carrés) [56] et la propriété statistique d'appariement des retards de corrélation d'ordres supérieurs [55]. Il atteint la borne de Cramér-Rao même pour des valeurs de RSB inférieures à 0 dB. L'estimateur proposé a été combiné à une technique itérative qui consiste à réduire par soustraction le bruit affectant le retard de corrélation d'ordre zéro. En plus de l'amplitude des coefficients de réflexion, cette technique considère l'utilisation de la valeur propre minimale de la matrice de corrélation comme condition d'arrêt prédéfinie. Même s'ils restent préliminaires, les résultats obtenus par l'utilisation de cet estimateur permettent de maintenir ouvertes des voies de recherche sur le sujet relativement peu exploré par les chercheurs.

Quant à la quatrième contribution, elle porte sur une étude empirique permettant de déterminer l'ordre qui conduit à l'obtention de la meilleure fidélité de l'enveloppe spectrale, d'un signal audio, à sa structure formantique, lorsque les deux traitements, paramétrique et réduction de bruit, sont combinés. Les résultats obtenus sont cohérents avec les travaux proposés dans la littérature dans la mesure où la stratégie de faire

précéder le traitement paramétrique classique par un traitement pour la réduction de bruit permet de réduire considérablement le bruit de fond tout en contournant les problèmes d'instabilité souvent rencontrés par les techniques classiques de modélisation linéaire prédictive.

1.4 Plan de la thèse

Ce rapport de thèse est structuré en 7 chapitres.

Le chapitre 2 rappelle les principes théoriques des techniques de modélisation paramétrique les plus communément utilisées, ces principes étant utiles pour les chapitres suivants. Le chapitre 3 présente une synthèse de la littérature portant sur les méthodes de traitement pour la réduction de bruit opérant dans les domaines de fréquences et de corrélation, jugée pertinente pour cette thèse.

Une méthode à deux microphones, particulièrement appropriée aux situations de bruit fortement non stationnaire et opérant dans le domaine fréquentiel, est présentée dans le chapitre 4. Considérée comme un prétraitement pour la modélisation paramétrique d'un signal audio, cette méthode est développée en détail dans un article de revue qu'on retrouve dans ce chapitre.

Le chapitre 5 présente deux versions d'une méthode itérative élaborée dans le domaine de corrélation et dédiée à l'ajustement de l'enveloppe spectrale d'un signal audio dégradé par la présence de bruit. Nous discutons à travers ce chapitre les

particularités qui différencient ces deux versions, lesquelles sont développées dans deux articles qu'on retrouve en annexe.

Le chapitre 6 est dédié aux points qui ont été étudiés, mais qui n'ont fait l'objet d'aucune publication. Ce chapitre aborde, dans un premier temps, le problème de détermination de l'ordre dans lequel les deux traitements (paramétrique et réduction de bruit) doivent être utilisés. Ensuite, il présente une nouvelle méthode que nous avons développée et qui permet d'améliorer la précision de l'estimateur de la variance du bruit dans le cas d'un signal audio contaminé par un bruit blanc additif. Quelques résultats préliminaires et intéressants sont inclus dans ce chapitre, ouvrant la voie à davantage de recherche sur le sujet.

Le chapitre 7 présente une discussion générale et conclut cette thèse.

CHAPITRE 2
RAPPELS

2.1 Introduction

L'analyse spectrale d'un signal audio discrétisé se base souvent sur des techniques de modélisation paramétrique, notamment la modélisation à moyenne mobile (MA), autorégressive (AR), et hybride autorégressive à moyenne mobile (ARMA). Ce sont des techniques récursives qui consistent à déterminer à partir des données observées un modèle rationnel linéaire qui représente au mieux le signal considéré. Elles se distinguent de celles non paramétriques (périodogramme et ses variantes) par deux points essentiels :

1. Elles permettent de réduire l'espace de représentation d'un signal audio en mémoire en conservant uniquement les paramètres du modèle. Ceci est particulièrement intéressant dans de nombreuses applications audio (transmission, classification, détection).
2. Elles permettent d'obtenir une très bonne résolution spectrale d'un signal audio d'une façon rapide (dès les premiers échantillons) et d'ajuster les paramètres du modèle point par point à partir des données observées.

Ce chapitre décrit brièvement les principes théoriques des trois modèles rationnels cités ci-dessus. Dans ce chapitre, nous nous intéressons plus particulièrement à la modélisation autorégressive (AR) qui est une méthode très populaire en analyse

spectrale paramétrique des signaux audio et par conséquent, choisie dans ce projet de thèse. Différentes méthodes usuelles d'estimation des paramètres d'un modèle AR seront également présentées. Dans ce qui suit, nous considérons un processus $x(n)$ stationnaire, de second ordre et centré.

2.2 Modèle à moyenne mobile « MA »

On dit que $x(n)$ est un processus aléatoire à moyenne mobile d'ordre q, noté MA(q), s'il peut s'écrire comme une combinaison linéaire de « q + 1 » valeurs de l'entrée $\varepsilon(n)$, supposée être un bruit blanc de moyenne nulle et de variance σ_ε^2 [1], c.-à-d.,

$$x(n) = \sum_{k=0}^{q} b_k \varepsilon(n-k) \tag{2.1}$$

Les coefficients $\{b_k\}$ constituent les paramètres du modèle. Désigné souvent par modèle « tout-zéro », un processus MA(q) peut être vu comme la sortie d'un filtre causal et linéaire excité par l'entrée $\varepsilon(n)$ et auquel est associée la transformée en Z, $B(z)$, donnée par :

$$B(z) = \sum_{k=0}^{q} b_k z^{-k} \tag{2.2}$$

La densité spectrale de puissance (DSP) d'un processus MA(q) est obtenue en évaluant le spectre en $z = e^{j\omega}$ [2], c.-à-d.,

$$S(e^{j\omega}) = \sigma_\varepsilon^2 \mid B(e^{j\omega}) \mid^2 \tag{2.3}$$

Contrairement à la transformée de Fourier (appliquée aux données observées), on a ainsi accès à une évaluation paramétrique du spectre.

2.3 Modèle autorégressif « AR »

On dit que $x(n)$ est un processus autorégressif d'ordre p, noté AR(p), s'il peut s'écrire comme une combinaison linéaire de ses « p » échantillons qui précèdent l'instant n, à un bruit blanc près [1], c.-à-d.,

$$\sum_{k=0}^{p} a_k x(n-k) = b_0 \varepsilon(n) \tag{2.4}$$

Les coefficients $\{a_k\}$ constituent les paramètres du modèle, alors que le processus $\{\varepsilon(n)\}$ représente un bruit blanc dont la variance σ_ε^2 est égale à l'erreur de prédiction (distance entre le modèle et les observations). En général, le terme b_0 est choisi de telle sorte que le premier coefficient a_0 soit égal à l'unité. Désigné souvent par modèle « tout-pôle », un processus AR(p) peut être vu comme la sortie d'un filtre causal et linéaire excité par l'entrée $\varepsilon(n)$ et auquel est associée la transformée en Z, $C(z)$, donnée par :

$$C(z) = \frac{b_0}{A(z)} = \frac{b_0}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{2.5}$$

La DSP d'un processus AR(p) est définie par [2] :

$$S(e^{j\omega}) = \frac{\sigma_\varepsilon^2 \mid b_0 \mid^2}{\mid A(e^{j\omega}) \mid^2} \tag{2.6}$$

Il convient de souligner que la stationnarité du processus AR(p) n'est garantie que si $C(z)$ est stable et causal, c.-à-d., les zéros du polynôme $A(z)$ sont situés à l'intérieur du cercle unité.

2.4 Modèle autorégressif à moyenne mobile « ARMA »

Obtenu par la combinaison des processus AR(p) et MA(q) et désigné souvent par modèle « zéro-pôle », un processus autorégressif à moyenne mobile d'ordre (p, q), noté ARMA(p, q), est défini par la relation de récurrence suivante [1] :

$$\sum_{k=0}^{p} a_k x(n-k) = \sum_{k=0}^{q} b_k \varepsilon(n-k) \tag{2.7}$$

Un processus ARMA(p, q) peut être vue comme la sortie d'un filtre causal et linéaire excité par l'entrée $\varepsilon(n)$ et auquel est associée la transformée en Z, $H(z)$, donnée par :

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^{q} b_k z^{-k}}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{2.8}$$

La DSP d'un processus ARMA(p, q) est définie par [2] :

$$S(e^{j\omega}) = \frac{\sigma_\varepsilon^2 \mid B(e^{j\omega}) \mid^2}{\mid A(e^{j\omega}) \mid^2} \tag{2.9}$$

Bien qu'il soit plus performant qu'un processus AR(p) en termes de résolution spectrale, un processus ARMA(p, q) s'avère plus coûteux en temps de calcul et ressources mémoire (dû notamment à l'évaluation dans un premier temps de la fonction de corrélation) [3].

2.5 Remarque sur le lien entre AR, MA et ARMA

Le théorème de décomposition de Wold [4] permet d'établir une équivalence entre les modèles ARMA, MA et AR. Essentiellement, ce théorème affirme qu'on peut approximer tout processus ARMA(p, q) stationnaire ou MA(q) de variance finie par un modèle AR unique d'ordre possiblement infini (c.-à-d., AR(∞)). De même, on peut approximer tout processus ARMA(p, q) ou AR(p) par un modèle MA d'ordre possiblement infini (c.-à-d., MA(∞)).

Ce théorème est important dans la mesure où il nous permet de choisir parmi ces trois modèles, un modèle non adéquat et néanmoins avoir une approximation raisonnable moyennant l'utilisation d'un ordre suffisamment élevé. Étant donné que l'évaluation des paramètres d'un modèle AR aboutit à un système d'équations linéaires, il a un avantage, en termes de temps de calcul, sur les techniques d'évaluation des paramètres des modèles ARMA et MA [5].

2.6 Évaluation des paramètres d'un processus AR(p)

Le problème d'estimation d'un modèle AR, souvent désignée par analyse « LPC », est équivalent à celui d'évaluation des coefficients d'un filtre « tout-pôle » excité par

une entrée inconnue et dont la sortie est connue. Les méthodes présentées ci-dessous sont couramment utilisées pour l'évaluation des paramètres d'un modèle AR.

2.6.1 Méthodes des moindres carrés

Dans ce type de méthodes, on cherche à minimiser, au sens des moindres carrés (LS pour « Least Squares ») et dans le modèle défini par l'équation (2.4), la puissance de l'erreur d'approximation (dite erreur de prédiction) d'un signal $x(n)$ [2] :

$$\min_{a_k}\left\{E_p = \sum_{n=n_0}^{n_1} |\,\varepsilon(n)\,|^2 = \sum_{n=n_0}^{n_1}\left|x(n)+\sum_{k=1}^{p} a_k x(n-k)\right|^2\right\} = \min_{a_k}\left\{\sum_{n=n_0}^{n_1}\left\|Xa+\mathrm{x}_1\right\|^2\right\} \qquad (2.10)$$

Rappelons que pour le cas $n_0 = 0$ et $n_1 = N$

$$X = \begin{bmatrix} x(0) & 0 & 0 & \cdots & 0 \\ x(1) & x(0) & 0 & \cdots & 0 \\ x(2) & x(1) & x(0) & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ x(p-1) & x(p-2) & x(p-3) & \cdots & x(0) \\ x(p) & x(p-1) & x(p-2) & \cdots & x(1) \\ \vdots & \vdots & \vdots & & \vdots \\ x(N-1) & x(N-2) & x(N-3) & \cdots & x(N-p) \\ x(N) & x(N-1) & x(N-2) & \cdots & x(N-p+1) \\ 0 & x(N) & x(N-1) & \cdots & x(N-p+2) \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & x(N) \end{bmatrix} : \quad \text{matrice de taille } (N+p)\times p,$$

$a = [a_1 \cdots a_p]^T$: vecteur paramètre du modèle de taille p, et

$\mathrm{x}_1 = [x(1) \quad x(2) \quad \cdots \quad x(N) \quad 0 \quad \cdots \quad 0]^T$: vecteur de taille $(N+p)$.

Le choix de l'intervalle $[n_0, n_1]$ sur lequel est effectuée la minimisation (intervalle d'analyse) conduit à des méthodes différentes, parmi lesquelles nous citons [6] :

- $[0, N]$ avec application d'une fenêtre rectangulaire de taille N ($n = 0,1,\cdots,N$) au signal $x(n)$: méthode de l'autocorrélation ;
- $[p, N]$ avec application d'une fenêtre rectangulaire de taille $N - p$ ($n = p, p+1\cdots,N$) au signal d'erreur : méthode de la covariance.

Indépendamment de la méthode utilisée, le vecteur $\hat{a}$ qui minimise l'équation (2.10) est donné par :

$$\hat{a} = -(X^T X)^{-1} X^T x_1 = -R_x^{-1} r \tag{2.11}$$

avec

$$R_x(i,j) = [X^T X](i,j) = \sum_{n=n_0}^{n_1} x^*(n-i)x(n-j) \tag{2.12}$$

$$r(j) = \sum_{n=n_0}^{n_1} x(n)x^*(n-j) \tag{2.13}$$

Dans ce qui suit, nous décrivons ces deux approches ainsi qu'une approche similaire connue sous le nom de la méthode du maximum d'entropie. Également, nous discutons brièvement des avantages et inconvénients de chacune d'elles.

2.6.2 Méthode de l'autocorrélation

Dans cette méthode, les paramètres d'un modèle AR(p) vérifient les équations normales de Yule-Walker qui unissent les paramètres et les retards de corrélation par le système d'équations linéaires suivant [3] :

$$r_x(n) = \begin{cases} r_x^*(-n), & n < 0 \\ -\sum_{k=1}^{p} a_k r_x(n-k) + \sigma_\varepsilon^2, & n = 0 \\ -\sum_{k=1}^{p} a_k r_x(n-k), & n \geq 1 \end{cases} \tag{2.14}$$

En général, l'évaluation des paramètres d'un modèle AR d'ordre p se fait en trois étapes :

a. choix de p équations pour $n > 0$ à partir du système d'équations (2.14);

b. estimation des paramètres $\{a_1, a_2, \ldots, a_p\}$ du modèle ;

c. estimation de σ_ε^2 pour $n = 0$ à partir du système d'équations (2.14).

En pratique, il convient de choisir le système d'équations qui nécessite le moins de retards de corrélation possible, c.-à-d., $n = 1,2,\cdots,p$. Ce système peut ainsi s'écrire sous forme matricielle de la façon suivante :

$$\begin{bmatrix} r_x(0) & r_x(-1) & \cdots & r_x(-p+1) \\ r_x(1) & r_x(0) & \cdots & r_x(-p+2) \\ \vdots & \vdots & \vdots\vdots & \vdots \\ r_x(p-1) & r_x(p-2) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = -\begin{bmatrix} r_x(1) \\ r_x(2) \\ \vdots \\ r_x(p) \end{bmatrix} \tag{2.15}$$

L'équation (2.15) peut s'écrire sous une forme plus compacte :

$$R_x a = -r_x \tag{2.16}$$

où R_x, a et r_x représentent respectivement la matrice d'autocorrélation, le vecteur des paramètres du modèle AR(p), et le vecteur d'autocorrélation.

En raison des propriétés de la matrice d'autocorrélation (Toeplitz et à symétrie hermitienne), le système d'équations (2.16) peut être résolu efficacement au moyen de l'algorithme de Levinson-Durbin [7] en $O(p^2)$ opérations. L'utilisation d'un estimateur avec biais des retards de corrélation dans (2.16), conduit toujours à un système à phase minimale (pôles à l'intérieur du cercle unité). Ceci n'est pas le cas en général pour un estimateur sans biais qui permet pourtant de fournir une meilleure estimation [8].

2.6.3 Méthode de la covariance

Dans cette méthode, les paramètres d'un modèle AR(p) vérifient le système d'équations linéaires suivant [8] :

$$\sum_{k=1}^{p} a_k \phi_x(i,k) = \phi_x(i,0), \qquad 1 \le i \le p \tag{2.17}$$

avec

$$\phi_x(i,k) = \sum_{n=p}^{N} x(n-k)x^*(n-i) \tag{2.18}$$

Ce système peut ainsi s'écrire sous forme matricielle de la façon suivante :

$$\begin{bmatrix} \phi_x(1,1) & \phi_x(1,2) & \cdots & \phi_x(1,p) \\ \phi_x(2,1) & \phi_x(2,2) & \cdots & \phi_x(2,p) \\ \vdots & \vdots & \vdots\vdots & \vdots \\ \phi_x(p,1) & \phi_x(p,2) & \cdots & \phi_x(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = -\begin{bmatrix} \phi_x(1,0) \\ \phi_x(2,0) \\ \vdots \\ \phi_x(p,0) \end{bmatrix} \tag{2.19}$$

Bien qu'hermitienne, la matrice de covariance $\phi_x(i,k)$ ne possède pas une structure Toeplitz. Ainsi, le système d'équations (2.19) ne peut pas être résolu par l'algorithme de Levinson-Durbin. Néanmoins, Morf [9] a élaboré un algorithme récursif requérant $O(p^2)$ opérations et permettant de résoudre ce système sans recourir à une inversion de matrice. Contrairement à la méthode de l'autocorrélation, la méthode de covariance ne nécessite aucune hypothèse sur les données à l'extérieur de l'intervalle d'analyse, permettant ainsi une analyse spectrale plus fine. Elle présente par contre l'inconvénient de ne pas garantir un système à phase minimale.

2.6.4 Méthode de Burg

Appelée souvent méthode du maximum d'entropie (MEM pour « Maximum Entropy Method »), on cherche dans cette méthode à minimiser la somme des erreurs de prédiction directe « forward » et rétrograde « backward » par rapport aux coefficients de réflexion Γ_j ($j = 1, \cdots, p$) [10] :

$$\min_{\Gamma_j} \left\{ E_j = \sum_{n=j}^{N} \left\{ |\varepsilon_j^+(n)|^2 + |\varepsilon_j^-(n)|^2 \right\} \right\} \tag{2.20}$$

avec

$$\varepsilon_j^+(n) = x(n) + \sum_{k=1}^{j} a_j(k)x(n-k) = \varepsilon_{j-1}^+(n) + \Gamma_j \varepsilon_{j-1}^-(n-1) \tag{2.21}$$

$$\varepsilon_j^-(n) = x(n-j) + \sum_{k=1}^{j} a_j^*(k)x(n-j+k) = \varepsilon_{j-1}^-(n-1) + \Gamma_j^* \varepsilon_{j-1}^+(n) \tag{2.22}$$

et où « + » et « - » désignent respectivement directe et rétrograde.

La solution à ce problème de minimisation est donnée par :

$$\Gamma_j = -\frac{2 \cdot \sum_{n=j}^{N} \varepsilon_{j-1}^+(n)[\varepsilon_{j-1}^-(n-1)]^*}{\sum_{n=j}^{N} \left\{ |\varepsilon_{j-1}^+(n)|^2 + |\varepsilon_{j-1}^-(n-1)|^2 \right\}} \tag{2.23}$$

Les paramètres d'un modèle AR(p) peuvent ainsi être déduits d'une façon récursive par l'algorithme de Levinson-Durbin. Bien que cette méthode garantisse un système à phase minimale ($|\Gamma_j| \leq 1$ pour tout j), elle présente les inconvénients de conduire à un dédoublement de raies spectrales dans le cas d'un signal sinusoïdal faiblement bruité, et à une sensibilité à la phase initiale du signal à modéliser [11], [12].

2.7 Critères de sélection de l'ordre d'un modèle AR(p)

Outre le problème de validation d'un modèle, le choix de son ordre est considéré comme un des problèmes fondamentaux en modélisation paramétrique :

- Choisir un ordre trop faible conduit à l'obtention d'un modèle qui ne représente pas les propriétés intrinsèques du signal (spectre lisse).

- Choisir un ordre trop élevé conduit à l'apparition dans le spectre de crêtes et de creux supplémentaires qui ne représentent pas nécessairement la structure formantique du signal.

De nombreux critères de décision ont été proposés dans la littérature dont le but est de faciliter la sélection de l'ordre d'un modèle paramétrique. Les plus connus sont : FPE (pour « Final Prediction Error »), MDL (pour « Minimum Description Length ») et AIC (pour « Akaike's Information Criterion ») [13]. Nous citons, par exemple, le critère FPE qui considère la plus petite valeur de p pour laquelle l'erreur de prédiction finale est minimale comme étant l'ordre du modèle. Bien qu'ils soient très utilisés, ces critères, heuristiques de nature, ont tendance à surévaluer l'ordre du modèle. Dans le cas d'un signal audio, le choix de l'ordre d'un modèle autorégressif est, en pratique, fonction du nombre de formants (une paire de pôles par formant) présents dans la bande passante du signal. Pour un signal audio échantillonné à la fréquence f_e (exprimée en Hz), l'ordre p est évalué selon la règle suivante [14] :

$$\frac{f_e}{1000} \leq p \leq \frac{f_e}{1000} + 4 \tag{2.24}$$

Ainsi, un ordre compris entre 11 et 15 est approprié pour la modélisation d'un signal audio échantillonné à 11.025 kHz.

2.8 Notion d'enveloppe spectrale

Le timbre d'un signal sonore est défini par l'ASA (pour « American Standard Association ») comme étant l'attribut perceptif permettant de distinguer deux sons

possédant la même hauteur tonale, la même intensité et la même durée. Un des indices perceptuels liés au timbre est l'enveloppe spectrale, qui représente la distribution en fréquences de l'énergie d'un signal. Ayant un rôle déterminant dans la perception des signaux audio, une estimation robuste de cette enveloppe contribue en effet à plus d'intelligibilité et de naturalité dans la voix. L'analyse spectrale autorégressive (analyse LPC) est particulièrement adaptée à la modélisation d'enveloppe spectrale d'un signal audio. La Figure 2.1 montre la superposition du spectre d'un signal de parole (échantillonné à 11.025 kHz) et différentes estimations d'enveloppes spectrales.

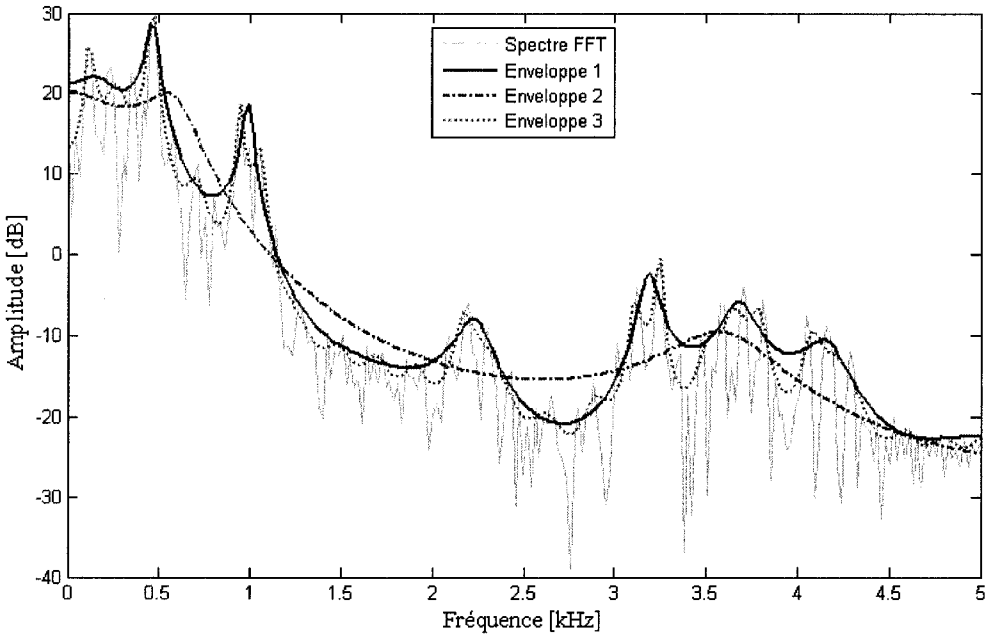

Figure 2.1 Spectre d'un signal de parole et enveloppes spectrales associées.

En passant par les crêtes du spectre, l'enveloppe 1 modélise convenablement la structure formantique du signal. Dans ce cas, l'ordre du modèle est adéquatement sélectionné. L'enveloppe 2 est caractérisée par un nombre très réduit de composantes

spectrales qui ne représentent pas la structure formantique avec suffisamment de précision. Ce problème surgit lorsque l'ordre du modèle est sous-estimé. Quant à l'enveloppe 3 qui est plus précise, elle met plutôt la structure fine du spectre en évidence. Dans ce cas, l'ordre du modèle paramétrique est surestimé. Rappelons que les formants vocaux sont liés aux fréquences de résonance du corps sonore de la voix, appelé conduit vocal.

CHAPITRE 3
REVUE DE LITTÉRATURE

Ce chapitre présente une revue de la littérature portant sur deux classes distinctes de méthodes de traitement des signaux audio pour la réduction de bruit. La première classe incorpore les méthodes qui opèrent dans le domaine fréquentiel. Dans cette classe, une attention particulière est accordée à deux approches, MS (pour « Minimum Statistics ») et Zelinski, sur lesquelles est basée notre méthode qui fera l'objet du chapitre suivant. Quant à la seconde classe, elle contient les méthodes élaborées dans le domaine de corrélation. Toutefois, peu de littérature est disponible au sujet de cette seconde classe de méthodes.

La plupart des méthodes de filtrage décrites dans ce chapitre considèrent que le signal audio est corrompu par un bruit de fond additif et indépendant de la source du signal. L'effet d'un bruit additif, associé acoustiquement (bruit ambiant) ou numériquement (bruit de quantification) à un signal audio peut conduire à la dégradation substantielle de son intelligibilité et de sa qualité de perception [15].

3.1 Méthodes élaborées dans le domaine fréquentiel

Les méthodes de cette classe peuvent être regroupées en deux catégories principales : méthodes à microphone unique et méthodes à plusieurs microphones.

Dans la première catégorie, on cherche à exploiter des informations statistiques sur le signal audio et le bruit. La soustraction spectrale et le filtrage de Wiener [16], [17] et

[18], font partie de cette catégorie. Bien que ces méthodes de filtrage soient efficaces à réduire le niveau de bruit de fond, elles souffrent de sérieuses limitations qui peuvent se résumer dans les trois points suivants. Premièrement, une sous-évaluation du niveau du bruit de fond pourrait résulter si les statistiques associées à son évolution changeaient rapidement dans le temps. Deuxièmement, le bruit résiduel généré comporte généralement des artefacts audibles, désignés dans la littérature par le terme de « bruit musical », qui sont désagréables pour l'oreille humaine. Troisièmement, une perte d'information utile (perte d'intelligibilité) pourrait avoir lieu si la puissance du signal de bruit était égale ou supérieure à la puissance du signal d'intérêt (RSB très faible).

Il convient de mentionner que la plupart de ces méthodes de filtrage nécessitent inévitablement un détecteur d'activité vocale VAD (pour « Voice Activity Detector ») qui permet d'identifier les périodes de silence dans un signal sonore. En effet, non seulement le raffinement d'un VAD est difficilement atteint, mais aussi son application à un signal sonore ayant un faible RSB génère souvent un phénomène d'écrêtage (distorsion).

Quant à la seconde catégorie, elle comprend les méthodes qui reposent sur la discrimination spatiale d'un réseau de microphones pour séparer le signal d'intérêt du bruit. Cette discrimination spatiale a été exploitée par Kaneda et Tohyama [19] qui ont développé un algorithme de mise en forme (« beamforming ») à deux microphones dans lequel le champ de bruit est considéré comme étant un espace non corrélé. Cette méthode a été étendue à un nombre arbitraire de microphones et combinée à un filtrage

de Wiener adaptatif par Zelinski [20], [21] pour améliorer davantage le signal à la sortie du réseau. Trois problèmes bien connus sont associés à l'utilisation de cette approche. Premièrement, les signaux de bruit issus des différents microphones contiennent souvent des composantes corrélées, particulièrement en basse fréquence, comme c'est le cas dans un champ acoustique diffus [22]. Deuxièmement, une telle approche génère un bruit résiduel audible ayant un spectre de puissance en forme de « cosinus » qui est perçu véritablement désagréable pour l'oreille humaine [23]. L'origine de ce bruit résiduel est notamment due à la substitution par sa partie réelle de la densité inter-spectrale de puissance (DISP) des signaux observés. Troisièmement, bien que l'application de la fonction de transfert de Zelinski à la sortie d'un réseau standard permet une réduction effective du bruit restant, elle génère en contrepartie un bruit résiduel (biais) proportionnel au nombre de microphones utilisés [24].

McCowan et Bourlard [25] ont remplacé l'hypothèse d'un champ de bruit non corrélé, considérée dans l'approche de Zelinski, par un modèle plus précis et plus général basé sur une connaissance préalable de la fonction de cohérence spatiale du champ de bruit. Cependant, les deux méthodes surestiment la DSP du bruit à la sortie du réseau et par conséquent, elles sont sous-optimales dans le sens de Wiener [24]. Pour remédier à ce problème de surestimation du bruit, Simmer et Wasiljeff [24] ont suggéré de substituer le dénominateur de la fonction de transfert de Zelinski par la DSP du signal à la sortie d'un réseau standard. Une version améliorée du postfiltre de McCowan a été proposée par Lefkimmiatis et Maragos [26]. Ils ont obtenu une évaluation plus précise de la DSP du signal à la sortie d'un réseau à réponse sans distorsion et à variance

minimale MVDR (« pour Minimum Variance Distortionless Response ») en prenant en compte la réduction de bruit acquise par ce dernier. Dans ce travail de thèse, nous proposons de combiner la méthode de Zelinski à un estimateur de la DSP du bruit afin de pallier les problèmes cités ci-dessus (chapitre 4).

La structure du suppresseur de lobes latéraux généralisé GSC (pour « Generalized Sidelobe Canceller ») a été considérée dans de nombreuses applications comme une mise en œuvre efficace d'un réseau adaptatif contraint par la direction. Initialement proposée par Griffiths et Jim [27], cette structure permet d'améliorer le RSB à la sortie d'un réseau standard sans introduire de nouvelles distorsions au signal estimé [28]. Néanmoins, la performance de cette structure en termes de réduction de bruit dépend amplement du degré de cohérence spatiale du champ de bruit. Pour pallier le problème des composantes spatialement incohérentes du bruit, Fisher and Simmer [29] ont proposé une méthode qui associe un GSC à un filtre de Wiener adaptatif permettant ainsi d'extraire les signaux issus de la direction d'observation.

Bitzer, Simmer et Kammeyer [30] ont examiné les limites théoriques en termes de réduction de bruit de la structure d'un GSC. Ils ont montré que cette structure est plutôt appropriée pour être utilisée dans des chambres anéchoïques et non dans un champ acoustique diffus. En utilisant un réseau de microphones large bande subdivisé en plusieurs sous-réseaux à directivité contrôlée, Fisher et Kammeyer [31] ont montré que la performance en termes de réduction de bruit de la structure résultante est quasiment indépendante des propriétés de corrélation du champ acoustique (c.-à-d., le système est

approprié pour un champ acoustique aussi bien diffus que cohérent). Cette structure en réseau a été examinée en détail par Marro, Mathieux et Simmer [32].

Cohen [33] a proposé d'incorporer dans un réseau de microphones à base d'une structure GSC, un postfiltre multicanaux particulièrement approprié pour être utilisé dans des environnements de bruit non stationnaire. Afin de discriminer les transitoires du signal d'intérêt de celles d'interférences, il a employé la sortie principale d'un réseau à base de GSC conjointement avec les signaux de bruit de référence. Pour permettre une implémentation en temps réel de la méthode, Cohen a suggéré dans un article subséquent [34] d'alimenter le réseau par les décisions de discrimination faites par le postfiltre. Bien que les systèmes à base de réseaux de microphones acquièrent une performance supérieure à celle obtenue au moyen d'un microphone unique, leur mise en oeuvre s'avère difficile et coûteuse en temps de calcul et ressources mémoire.

En considérant une implémentation facile et avantageuse en temps de calcul et ressources mémoire, les systèmes à base d'une paire de microphones semblent être intéressants pour les deux raisons suivantes : ils garantissent une performance suffisamment acceptable, en particulier pour les applications portables et compactes (les prothèses auditives numériques et les téléphones à mains libres) et ils sont fiables en termes de consommation d'énergie.

Dans ce contexte, Le Bouquin-Jeannes, Azirani et Faucon [23] ont proposé de combiner les fonctions de transfert relatives au filtre de Wiener et celui basé sur la fonction de cohérence à un estimateur de DISP afin de prendre la présence de quelques

composantes corrélées du bruit de fond en considération. Dans cette approche, la DISP des deux signaux observés a été moyennée sur les périodes de silence et soustraite de celle obtenue en présence d'une activité sonore. Guerin, Le Bouquin-Jeannes et Faucon [28] ont suggéré un estimateur de paramètre de lissage adaptatif permettant de déterminer la DISP du bruit qui devrait être utilisée dans la fonction de transfert d'un filtre à base de la fonction de cohérence. En évaluant la surestimation requise pour la DISP du bruit, ils ont montré que le bruit musical (résultant des larges fluctuations du paramètre de lissage entre les périodes d'activité sonore et celles de silences) peut être soigneusement contrôlé, particulièrement pendant les périodes d'activité sonore.

Zhang et Jia [35] ont proposé une simple procédure de décision quantifiée basée sur l'algorithme « MS » pour évaluer la DISP du bruit. Bien que les méthodes à base de la soustraction inter-spectrale (SIS) garantissent une performance satisfaisante dans une variété de bruits environnants, elles sont souvent inappropriées à traiter convenablement les bruits fortement non stationnaires, tels que le bruit impulsif et le bruit d'ambiance de type « cocktail-party ».

En situation de bruit fortement non stationnaire, le spectre du bruit de fond doit être évalué et mit à jour fréquemment, permettant ainsi une réduction effective du bruit. Au cours de la dernière décennie, de nombreuses méthodes ont été proposées dans le but d'évaluer continuellement le spectre du bruit de fond sans avoir recours à une détection explicite des périodes d'activité vocale. Martin [36] a proposé une méthode statistique, connue sous le nom de méthode de « MS », permettant d'évaluer le spectre du bruit de

fond en poursuivant les minima de la densité spectrale de puissance (DSP) du signal observé sur une fenêtre coulissante de largeur finie et pouvant contenir des segments audio de puissance assez élevée. Rappelons que cette méthode est basée sur deux hypothèses essentielles. Premièrement, les signaux d'intérêt et de bruit sont considérés comme étant deux processus aléatoires statistiquement indépendants. Deuxièmement, le niveau d'énergie d'un signal corrompu est fréquemment réduit à des valeurs représentatives du niveau d'énergie du bruit même dans les périodes d'activité sonore. Cette situation pourrait survenir dans les périodes de silence ou dans de brèves périodes entre les mots et les syllabes.

Basé sur les mêmes hypothèses que la méthode de « MS », l'algorithme récursif à minima contrôlés MCRA (pour « Minima Controlled Recursive Averaging ») proposé par Cohen [37] permet d'évaluer le spectre du bruit de fond en poursuivant les périodes de silence dans le signal observé (corrompu). Ces périodes sont obtenues en comparant le rapport entre l'énergie locale du signal observé à son minimum local contre un seuil fixe. Dans une version améliorée de la méthode MCRA [38], une approche différente, basée sur une évaluation de la probabilité de présence d'activité sonore, a été utilisée pour poursuivre les périodes de silence dans le signal observé.

En raison notamment de sa simplicité de mise en œuvre en temps réel, la méthode de « MS » a été choisie dans ce projet de thèse pour évaluer la DSP du bruit de fond.

3.2 Méthodes élaborées dans le domaine de corrélation

Les méthodes de cette classe peuvent également être regroupées en deux catégories principales : méthodes d'estimation paramétrique basées sur le modèle ARMA et méthodes de compensation paramétrique basées sur les équations normales de Yule-Walker d'ordres inférieurs (LOYWE pour « Low-Order Yule-Walker Equations »).

Dans la première catégorie, on retrouve les méthodes classiques qui reposent sur les équations de Yule-Walker d'ordres supérieurs (HOYWE pour « High-Order Yule-Walker Equations ») [5] et [39], la méthode basée sur un système d'équations normales surdéterminées (ODNE pour « Overdetermined Normal Equations ») [1], la méthode du maximum de vraisemblance (MLSE pour « Maximum Likelihood Spectral Estimation ») [40], et la méthode de l'erreur de prédiction récursive (RPE pour « Recursive Prediction Error ») de Gauss-Newton [41].

Les approches HOYWE considèrent l'hypothèse que dans le domaine de corrélation seul le retard de corrélation d'ordre zéro est affecté par un bruit blanc additif, alors que les retards de corrélation d'ordres supérieurs sont maintenus inchangés [42]. L'avantage de ces approches réside essentiellement dans leur mise en œuvre qui ne dépend pas du retard de corrélation d'ordre zéro. Ainsi, une évaluation sans biais des paramètres spectraux du modèle ARMA d'un processus AR peut être obtenue en présence d'un bruit blanc. Elles présentent par contre l'inconvénient de ne pas fournir une estimation robuste des retards de corrélation d'ordres supérieurs. Pour compenser les erreurs d'estimation générées par ces approches, la méthode ODNE, initialement proposée par

Cadzow [1], est souvent considérée. Plutôt que de se borner à l'ordre du modèle de prédiction, cette méthode considère l'utilisation d'un nombre assez élevé de corrélations pour obtenir les équations normales. Ce système d'équations normales surdéterminées est ensuite résolu au sens des moindres carrés. Il a été démontré dans [43] que la méthode ODNE fournit des paramètres spectraux robustes et stables.

Basée sur une procédure itérative d'optimisation non linéaire, la méthode MLSE cherche à trouver les paramètres d'un modèle ARMA qui maximisent une fonction de vraisemblance conditionnelle (produit des densités de probabilité individuelles de la séquence de données observées). Bien que cette méthode soit statistiquement efficace (elle converge souvent vers un optimum), sa mise en œuvre s'avère difficile et coûteuse en temps de calcul et ressources mémoire [44]. La méthode RPE repose sur des informations connues a priori sur le signal (propriétés spectrales du signal en présence de bruit, niveau de distorsion spatiale) pour évaluer d'une manière adaptative les paramètres d'un modèle ARMA sous contraintes. Il est bien connu que cette méthode est plus fiable et avantageuse en temps de calcul et ressources mémoire que la méthode MLSE [41].

Quant à la seconde catégorie, elle regroupe les méthodes qui utilisent les LOYWE pour compenser les effets d'un bruit de fond sur les paramètres spectraux d'un processus AR. Pour ces méthodes, la stabilité des paramètres spectraux estimés est un préalable essentiel pour la validité du modèle paramétrique choisi. Dans [45], un algorithme de recherche dichotomique est utilisé dans le but d'obtenir un biais approprié qui devrait

être soustrait du retard de corrélation d'ordre zéro sans compromettre la propriété de la matrice de corrélation d'être définie positive. La compensation des effets d'un bruit de fond dans [46] est obtenue en soustrayant itérativement de la séquence de corrélation du signal observé une estimation de la puissance du bruit. Dans cette méthode, la puissance du bruit est supposée être connue. Enfin, il convient de préciser que les méthodes proposées dans le chapitre 5 appartiennent à cette seconde catégorie.

CHAPITRE 4
RÉDUCTION DE BRUIT DANS LE DOMAINE FRÉQUENTIEL

4.1 Résumé

Ce chapitre décrit une méthode élaborée dans le domaine fréquentiel que nous avons développée dans le but de réduire le bruit de fond présent dans un signal audio large bande. Il s'agit d'une méthode à deux microphones qui repose sur une technique de filtrage adaptatif proposée par Zelinski [20] et un estimateur de la densité spectrale de puissance (DSP) du bruit, désigné par algorithme de MS (voir section 5.1) et introduit par Martin [36].

Dans un premier temps, nous avons réalisé une étude de la technique de filtrage adaptatif, qui nous a permis d'identifier quelques problèmes inhérents associés à son utilisation, notamment la génération d'un bruit résiduel audible et déplaisant à l'oreille humaine, et l'inefficacité à réduire le bruit cohérent inter-canaux. Ensuite et dans un souci de pallier ces problèmes, nous proposons de modifier l'estimateur de la DSP du bruit introduit par Martin et le combiner à cette méthode de filtrage au moyen d'une procédure de décision quantifiée. Cette combinaison possède deux fonctions essentielles : rendre inaudible le bruit résiduel généré par le filtrage adaptatif, et réduire le bruit cohérent (particulièrement en basse fréquence) nécessairement présent dans un champ acoustique lointain. Enfin, une structure permettant que le filtrage adaptatif et l'estimateur de la DSP du bruit soient réalisés conjointement dans le domaine fréquentiel

est proposée, permettant ainsi un bon compromis entre la réduction de bruit et la distorsion produite sur le signal d'intérêt. Cette méthode a fait l'objet d'un rapport technique [47], d'un article de conférence [48] et d'un article de revue [49]. Seule la version retenue de l'article de revue est présentée dans ce chapitre.

Les contributions principales de cet article sont résumées dans les points suivants :

- Élaborée dans le domaine fréquentiel, cette méthode permet d'obtenir un bon compromis entre la réduction de bruit et la distorsion produite sur le signal d'intérêt.
- La méthode proposée est particulièrement appropriée aux bruits de fond fortement non stationnaires (bruit impulsif et bruit d'ambiance de type « cocktail-party ») qui sont difficilement traités par les méthodes concurrentes.
- La complexité calculatoire de la méthode est approximativement de $O(N\log(N)+N+D)$. Cette complexité calculatoire relativement faible permet d'envisager une implémentation en temps réel de la méthode.

- La méthode est considérée avantageuse en temps de calcul et ressources mémoire.

Le contenu intégral de l'article [49] est présenté dans les pages suivantes. Ce contenu correspond à celui pour la parution dans *IEEE Transactions on Audio, Speech and Language Processing* (soumission en nov. 2008).

4.2 A Two-Microphone Algorithm for Speech Enhancement

Abstract–In this paper, we focus on the problem of enhancing a speech signal contaminated with additive noise when noisy observations from two microphones are available. An existing approach proposed by Zelinski is known to have two shortcomings. First, the method lacks robustness in a number of practical noise fields (i.e., coherent noise). Second, it gives rise to a residual noise that is unpleasant to human listeners. To overcome these drawbacks, we propose to modify this method by incorporating an appropriate noise power spectrum estimator. Based on minimum statistics and a soft-decision scheme, this estimator seeks to provide a good tradeoff between noise reduction and speech distortion. We call the proposed method the modified Zelinski approach (MZA). Analysis of objective measures and speech spectrograms, as well as subjective listening tests, show that the proposed method outperforms the cross-spectral subtraction (CSS) approach, in particular for highly nonstationary noise.

4.3 Introduction

In various applications such as, mobile communications and digital hearing aids, the presence of interfering noise may cause serious deterioration in the perceived quality of speech signals. Thus, there exists considerable interest in developing speech enhancement algorithms that solve the problem of noise reduction in order to make the compensated speech more pleasant to a human listener. Noise reduction problem in single and multiple microphone environments has been extensively studied [1]. Single

microphone speech enhancement approaches often fail to yield satisfactory performance, in particular when the interfering noise statistics are time-varying [2]. In contrast, multiple microphone systems with more than two microphones (i.e., Cohen post-filter [3]) provide superior performance over the single microphone schemes at the expense of a substantial increase of implementation complexity and computational cost.

In this paper, we address the problem of enhancing a speech signal corrupted with additive noise when observations from two microphones are available. Considering ease of implementation and lower computational cost when compared with approaches requiring microphone arrays with more than two microphones, two-microphone solutions are yet a promising class of speech enhancement systems due to their simpler array processing, which is expected to lead to lower power consumption, while still maintaining sufficiently good performance, in particular for compact portable applications (i.e., digital hearing aids, and hands-free telephones). The adaptive noise canceller (ANC) [4], [5], and cross-spectral subtraction (CSS) [6]–[8] are well-known examples. The standard ANC method provides high speech distortion in the presence of any crosstalk interferences between the two microphones. Widely reported in the literature, the CSS-based approach provides interesting performance in a variety of noise fields. However, it lacks efficiency in dealing with highly nonstationary noises such as the multitalker babble.

To deal with these limitations, we propose a method that combines the existing Zelinski's approach, in the case of two-microphone arrangement, with an appropriate

noise power spectrum estimator, which is particularly suitable for highly nonstationary noise environments. Based on minimum statistics and a soft-decision scheme, this estimator seeks to provide a good tradeoff between the amount of noise reduction and the speech distortion, while attenuating the high energy correlated noise components (i.e., coherent direct path noise), especially in the low frequency ranges.

The rest of the paper is organized as follows. Section II surveys the state of the art in speech enhancement. In Section III, we review Zelinski's approach in the case of two-microphone arrangement. In Section IV, we describe the single channel noise spectrum estimation algorithm used to cope with Zelinski's approach shortcomings, and use this algorithm in conjunction with a soft-decision scheme to come up with the proposed method. Section V provides objective measures, speech spectrograms and subjective listening test results from experiments comparing the performance of the proposed method (MZA) with the CSS-based approach. Finally, Section VI concludes the paper.

4.4 State of the art

There have been several approaches proposed to deal with the noise reduction problem in speech processing, with varying degrees of success. These approaches can generally be divided into two main categories. The first category uses a single microphone system and exploits information about the speech and noise signal statistics for enhancement. The most often used single microphone noise reduction approaches are the spectral subtraction method and its variants [9].

The second category of signal processing methods applicable to that situation involves using a microphone array system. These methods take advantage of the spatial discrimination of an array to separate speech from noise. This spatial information has been exploited by Kaneda and Tohyama [10] who developed a two-microphone beamforming algorithm which considers spatially uncorrelated noise field. This method has been extended to an arbitrary number of microphones and combined with an adaptive Wiener filtering by Zelinski [11], [12] to further improve the output of the beamformer.

McCowan and Bourlard have replaced the spatially uncorrelated noise field assumption by a more accurate model based on an assumed knowledge of the noise field coherence function, and extended Zelinski's approach to develop a more appropriate post-filtering scheme [13]. However, both methods overestimate the noise power density at the beamformer's output and, therefore, are suboptimal in the Wiener sense [14]. An improved version of the existing McCowan post-filter has been proposed by Lefkimmiatis and P. Maragos [15]. They have obtained a more accurate estimation of the noise power spectral density at the output of the beamformer by taking into account the noise reduction performed by the minimum variance distortionless response (MVDR) beamformer.

The generalized sidelobe canceller (GSC) method, initially introduced by Griffiths and Jim [16], has been considered for the implementation of adaptive beamformers in various applications. It has been found that this method performs well in enhancing the

signal-to-noise ratio (SNR) at the beamformer's output without introducing further distortion to the desired signal components [7]. However, the achievable noise reduction performance is limited by the amount of incoherent noise. To cope with the spatially incoherent noise components, Fisher and Simmer proposed a GSC based method which incorporates an adaptive Wiener filter in the look direction, [17]. Bitzer *et al.* have investigated the theoretical noise reduction limits of the GSC [18]. They have shown that this structure performs well in anechoic rooms, but it does not work well in diffuse noise fields. By using a broadband array beamformer partitioned into several harmonically nested linear subarrays, Fisher and Kamayer [19] have shown that the resulting noise reduction system performance is nearly independent of the correlation properties of the noise field (i.e., the system is suitable for diffuse as well as for coherent noise field). This array structure has been investigated in detail by Marro *et al.* [20].

Cohen [3] proposed to incorporate into the GSC beamformer a multichannel postfilter which is appropriate to work in nonstationary noise environments. To discriminate desired speech transients from interfering transients, he used both the GSC beamformer primary output and the reference noise signals. To get a real time implementation of their method, Cohen *et al.* suggested in a later paper [21] feeding back to the beamformer the discrimination decisions made by the postfilter.

In the two-microphone noise reduction context, Le Bouquin-Jeannès *et al.* [6] have proposed to modify both the Wiener and the coherence-magnitude based filters by including a cross power spectrum estimation to take some correlated noise components

into account. With that method, the cross power spectral density (cross-PSD) of the two input signals was averaged during speech pauses and subtracted from the estimated cross-PSD in the presence of speech. Guerin *et al.* [7] suggested an adaptive smoothing parameter estimator to determine the noise cross-PSD that should be used in the coherence-magnitude based filter. By evaluating the required overestimation for the noise cross-PSD, they showed that the musical noise (resulting from large fluctuations of the smoothing parameter between speech and non-speech periods) can be carefully controlled, especially during speech activity. Zhang and Jia [8] proposed a simple soft-decision scheme based on minimum statistics to estimate the noise cross-PSD.

4.5 Zelinski's approach in the case of two-microphone arrangement

This section introduces the signal model and gives a brief review of Zelinski's approach in the case of two-microphone arrangement. Let $s(t)$ be a speech signal of interest, and let the signal vector $n(t) = [n_1(t)\ n_2(t)]^T$ denote two-channel noise signals at the output of 2 spatially separated microphones. The sampled noisy signal $x_m(i)$ observed at the mth microphone can then be modeled as

$$x_m(i) = s(i) + n_m(i), \quad m = 1, 2 \tag{4.1}$$

where i is the sampling time index. The observed noisy signals are segmented into overlapping time frames by applying a window function and they are transformed into

the frequency domain using the short-time Fourier transform (STFT). Thus, we have for a given time frame :

$$X(k,l) = S(k,l) + N(k,l) \tag{4.2a}$$

where k is the frequency bin index, and l is the time index, and where

$$X(k,l) = [X_1(k,l)\ \ X_2(k,l)]^T \tag{4.2b}$$

$$N(k,l) = [N_1(k,l)\ \ N_2(k,l)]^T \tag{4.2c}$$

Zelinski's noise reduction system is derived from Wiener's theory, which solves the problem of optimal signal estimation in the mean-square error sense [12]. The Wiener filter weights the spectral components of the noisy signal according to the signal-to-noise power density ratio at individual frequencies given by :

$$W(k,l) = \frac{\Phi_{SS}(k,l)}{\Phi_{X_m X_m}(k,l)} \tag{4.3}$$

where $\Phi_{SS}(k,l)$ and $\Phi_{X_m X_m}(k,l)$ are respectively the power spectral densities (PSDs) of the desired signal and the input signal to the mth microphone.

For the formulation of Zelinski's approach, the following assumptions are made :

1. The noise signals are spatially uncorrelated, $E\{N_1^*(k,l) \cdot N_2(k,l)\} = 0$;
2. The desired signal $S(k,l)$ and the noise signal $N_m(k,l)$ are statistically independent random processes, $E\{S^*(k,l) \cdot N_m(k,l)\} = 0,\ m = 1,2$;
3. The noise PSDs are the same on the two microphones.

Under those assumptions, the unknown PSD $\Phi_{SS}(k,l)$ in (4.3) can be obtained from the estimated spatial cross-PSD $\Phi_{X_1X_2}(k,l)$ between microphone noisy signals. To improve the estimation, the estimated PSDs are averaged over the microphone pair, leading to the following transfer function :

$$\hat{W}(k,l) = \frac{\Re\{\hat{\Phi}_{X_1X_2}(k,l)\}}{(\hat{\Phi}_{X_1X_1}(k,l) + \hat{\Phi}_{X_2X_2}(k,l))/2} \tag{4.4}$$

where $\Re\{\cdot\}$ is the real operator, and "$\hat{}$" denotes an estimated value. It should be noted that only the real part of the estimated cross-PSD in the numerator of equation (4.4) is used, based on the fact that both the auto power density of the speech signal and the spatial cross power density of a diffuse noise field are real functions.

There are three well known problems associated with the use of Zelinski's approach. First, the noise signals on different microphones frequently hold correlated components, especially in low frequency ranges, as is the case in a diffuse noise field [22]. Second, such approach usually gives rise to an audible residual noise that has a cosine shaped power spectrum that is not pleasant to a human listener [6]. Third, applying Zelinski's transfer function to the output signal of a conventional beamformer yields an effective reduction of the remaining noise components but at the expense of an increased noise bias, especially when the number of microphones is large [14].

In this paper, we will focus our attention on estimating and discarding the residual and coherent noise components resulting from the use of Zelinski's approach in the case

of two-microphone arrangement. For such system, the overestimation of the noise power density should not be a problem.

4.6 Two-microphone speech enhancement system

In this section, we review the basic concepts of the existing noise spectrum estimation algorithm used to cope with Zelinski's approach shortcomings. Then, we use a variation of that algorithm in conjunction with a soft-decision scheme to come up with the proposed two-microphone processing method.

4.6.1 Noise Power Spectrum Estimation

For highly nonstationary environments (such as the multitalker babble), the noise spectrum needs to be estimated and updated continuously to allow an effective noise reduction. A variety of methods have been recently reported that continuously update the noise spectrum estimate while avoiding the need for explicit speech pause detection. Martin [23] proposed a method, known as the minimum statistics (MS), for estimating the noise spectrum based on tracking the minimum of the noisy speech over a finite window. Cohen [24] suggested a minima controlled recursive algorithm (MRCA) which updates the noise spectrum estimate by tracking the noise-only periods of the noisy speech. These periods are found by comparing the ratio of the noisy speech to the local minimum against a fixed threshold. In the improved MRCA approach [25], a different method was used to track the noise-only periods of the noisy signal based on the estimated speech-presence probability. Because of its simplicity that facilitates

affordable (hardware, power and energy wise) real-time implementation, the MS method was chosen in this paper for estimating the noise power spectrum.

The MS algorithm tracks the minima of a short term power estimate of the noisy signal within a time window of about 1 s. Let $\hat{P}(k,l)$ denotes the smoothed spectrum of the noisy signal $X(k,l)$, estimated at frequency k and frame l according to the following first-order recursive averaging

$$\hat{P}(k,l) = \hat{\alpha}(k,l) \cdot \hat{P}(k,l-1) + (1-\hat{\alpha}(k,l)) \cdot |X(k,l)|^2 \tag{4.5}$$

where $\hat{\alpha}(k,l)$ $(0 < \hat{\alpha}(k,l) < 1)$ is a time and frequency dependent smoothing parameter. The spectral minimum at each time and frequency index is obtained by tracking the minimum of D successive estimates of $\hat{P}(k,l)$, regardless of whether speech is present or not

$$\hat{P}_{\min}(k,l) = \min(\hat{P}_{\min}(k,l-1), \hat{P}(k,l)) \tag{4.6}$$

Because the minimum value of a set of random variables is smaller than their average, the noise spectrum estimate is usually biased. Let $B_{\min}(k,l)$ denotes the factor by which the minimum is smaller than the mean. This bias compensation factor is determined as a function of the minimum search window length D and the inverse normalized variance $Q_{eq}(k,l)$ of the smoothed spectrum estimate $\hat{P}(k,l)$. The resulting unbiased estimator of the noise spectrum $\hat{\sigma}_n^2(k,l)$ is then given by

$$\hat{\sigma}_n^2(k,l) = B_{\min}(k,l) \cdot \hat{P}_{\min}(k,l) \tag{4.7}$$

To make the adaptation of the minimum estimate faster, the search window of D samples is subdivided into U subwindows of V samples ($D = U \cdot V$) and the noise PSD estimate is updated every V subsequent PSD estimates $\hat{P}(k,l)$. In case of a sudden increase in the noise floor, the noise PSD estimate is updated when a local minimum with amplitude in the vicinity of the overall minimum is detected. The minimum estimate, however, lags by at most $D + V$ when the noise power increases abruptly. It should be noted that Martin's noise power estimator tends to underestimate the noise power, in particular when frame-wise processing with considerable frame overlap is performed. This underestimation problem is known and further investigation on the computation and correction of the bias of the minimum can be found in [26] and [27].

4.6.2 Proposed Algorithm

Although Zelinski's noise reduction method has shown its effectiveness in various practical noise fields, its performance could be increased if the residual and coherent noise components were estimated and discarded from the output spectrum. In the proposed two-microphone noise reduction method, this is done by using a variation of the MS method, in conjunction with a soft-decision scheme, in order to provide a good tradeoff between noise reduction and speech distortion. Fig. 1 shows an overview of the proposed two-microphone algorithm, which is described in details in this section.

We consider the case in which the STFT average of the noisy observations received by the two microphones, $Y(k,l)=(X_1(k,l)+X_2(k,l))/2$, is multiplied by a spectral gain function $G(k,l)$ for approximating the sound signal of interest, i.e.,

$$\hat{S}(k,l)=G(k,l)\cdot Y(k,l) \tag{4.8}$$

The gain function $G(k,l)$ is obtained by using equation (4.4), and can be expressed in the following extended form as

$$G(k,l)=\frac{(|X_1(k,l)|\cdot|X_2(k,l)|)\cdot\cos(\Delta\varphi(k,l))}{(|X_1(k,l)|^2+|X_2(k,l)|^2)/2} \tag{4.9a}$$

where

$$\Delta\varphi(k,l)=\varphi_{X_1}(k,l)-\varphi_{X_2}(k,l) \tag{4.9b}$$

and where $\varphi_{X_1}(k,l)$ and $\varphi_{X_2}(k,l)$ denote the phase spectra of the STFTs of $X_1(k,l)$ and $X_2(k,l)$ respectively that satisfy the condition $|\varphi_{X_1}(k,l)-\varphi_{X_2}(k,l)|<\pi/2$. In our implementation, any negative values of the gain function $G(k,l)$ were reset to a minimum spectral floor, on the assumption that such frequencies cannot be recovered. Moreover, we have obtained good results when the gain function $G(k,l)$ was squared, which improves signals selectivity (i.e., those coming from the direct path).

To track the residual and coherent noise components that are often present in the estimated spectrum in (4.8), a variation of the MS algorithm is implemented as follows. In performing the running spectral minima search, the D subsequent noise PSD

estimates are divided into 2 sliding data subwindows of $D/2$ samples. Whenever $D/2$ samples are processed, the minimum of the current subwindow is stored for later use. The sub-band noise power estimate $\hat{\sigma}_n^2(k,l)$ is obtained by picking the minimum value of the current signal PSD estimate and the latest $D/2$ PSD values. The sub-band noise power is updated at each time step. As a result, a fast update of the minimum estimate is achieved in response to a falling noise power. In case of a rising noise power, the update of the minimum estimate is delayed by D samples.

For accurate power estimates, the bias correction factor introduced by Martin was adjusted (i.e., scaled) by a constant decided empirically. This constant was obtained by performing the MS algorithm on a white noise signal so that the estimated output power had to match exactly that of the driving noise in the mean sense.

To discard the estimated residual and coherent noise components, a soft-decision scheme is implemented. For each frequency bin k and frame index l, we estimate the signal to noise ratio. The signal power is estimated from equation (4.8) and the noise power is the latest estimated value from equation (4.7). This ratio, called difference in level (DL), is calculated as follows

$$DL = 10 \cdot \log_{10}\left(\frac{|\hat{S}(k,l)|^2}{\hat{\sigma}_n^2(k,l)}\right) \tag{4.10}$$

The estimated DL value is then compared to a fixed threshold Th_s decided empirically. Based on that comparison, a running decision is taken by preserving the

sound frequency bins of interest and reducing the noise bins to a minimum spectral floor. That is

$$|\hat{\hat{S}}(k,l)|=\begin{cases}|\tilde{S}(k,l)|\cdot\lambda, & \text{if } DL<0\\ |\tilde{S}(k,l)|\cdot\left(\left(\frac{DL}{Th_s}\right)^2\cdot(1-\lambda)+\lambda\right), & \text{if } DL<Th_s\\ |\tilde{S}(k,l)|, & \text{otherwise.}\end{cases} \tag{4.11a}$$

where

$$|\tilde{S}(k,l)|=\sqrt{|\hat{S}(k,l)|^2-\hat{\sigma}_n^2(k,l)} \tag{4.11b}$$

and where λ is chosen such that $20\cdot\log_{10}(\lambda)=-40$ dB. When the estimated DL value is lower than the statistical threshold, the quadratic function "$(DL/Th_s)^2\cdot\cdot(1-\lambda)+\lambda$" allows the estimated spectrum to be smoothed during noise reduction. It should be noted that the so called DL has to take positive values during speech activity and negative values during speech pause periods.

Finally, the estimated magnitude spectrum in (4.11) was combined with the phase spectrum average of the two received signals prior to estimating the time signal of interest. In addition to the 6 dB reduction in phase noise, the time waveform resulting from such combination does offer a better match of the sound signal of interest coming from the direct path. After an inverse DFT of the enhanced spectrum, the resultant time waveform is half-overlapped and added to adjacent processed segments to produce an approximation of the sound signal of interest (Figure 4.1).

4.7 Performance Evaluation and Results

This section presents the performance evaluation of the proposed method (MZA), as well as the results of experiments comparing our method with the CSS approach. In all the experiments, the analysis frame length was set to 1024 data samples (23 ms at 44.1 kHz sampling rate) with 50% overlap. The analysis and synthesis windows thus had a perfect reconstruction property. The sliding window length of D subsequent PSD estimates was set to 100 samples. The threshold Th_s was fixed to 5 dB. The recordings were made using a Presonus Firepod recording interface and two Shure KSM137 cardioid microphones placed approximately 20cm apart. The experimental environment of the proposed two-microphone system is depicted in Figure 4.2. The room with dimensions of 5.5 x 3.5 x 3 m enclosed a speech source situated at a distance of 0.5 m directly in front (0 degrees azimuth) of the input microphones, and a masking source of noise located at a distance of 0.5 m from the speech source.

Designed to be equally intelligible in noise, five sentences taken from the Hearing in Noise Test (HINT) database [28] were recorded at a sampling frequency of 44.1 kHz.

- Sentence 1 (male talker): "Flowers grow in the garden".
- Sentence 2 (female talker): "She looked in her mirror".
- Sentence 3 (male talker): "The shop closes for lunch".
- Sentence 4 (female talker): "The police helped the driver".
- Sentence 5 (male talker): "A boy ran down the path".

Four different noise types, namely white Gaussian noise, helicopter rotor noise, impulsive noise and multitalker babble noise, were recorded at the same sampling rate and used throughout the experiments. The noise was scaled in power level and added acoustically to the above sentences with a varying SNR. A global SNR estimation of the input data was used. It was computed by averaging power over the whole length of the two observed signals with :

$$\mathrm{SNR} = 10 \cdot \log_{10} \left(\frac{\sum_{m=1}^{2} \sum_{i=1}^{I} s^2(i)}{\sum_{m=1}^{2} \sum_{i=1}^{I} (x_m(i) - s(i))^2} \right) \tag{4.12}$$

where I is the number of data samples of the signal observed at the mth microphone. Throughout the experiments, we used the average of the two clean signals $s(i) = (s_1(i) + s_2(i))/2$ as the clean speech signal. Objective measures, speech spectrograms and subjective listening tests were used to demonstrate the performance improvement achieved with the proposed method over the alternative approach.

4.7.1 Objective Measures

The Itakura-Saito (IS) distance [29] and the log spectral distortion (LSD) [30] were chosen to measure the differences between the clean and the test spectra. The IS distance has a correlation of 0.59 with subjective quality measures [31]. A typical range for the IS distance is 0–10, where lower values indicate better speech quality. The LSD provides reasonable degree of correlation with subjective results. A range of 0–15 dB can be considered for the selected LSD, where the minimum value of LSD corresponds to the

best speech quality. In addition to the IS and LSD measures, a frame-based segmental SNR was used which takes into consideration both speech distortion and noise reduction. In order to compute these measures, an utterance of the sentence 1 was processed through the two methods. The input SNR was varied from –8 dB to 8 dB in 4 dB steps.

Values of the IS distance measure for various noise types and different input SNRs are presented in Table I for signals processed by the different methods. Results in this table were obtained by averaging the IS distance values over the length of sentence 1. The results in this table indicate that the CSS approach yields more speech distortion than that produced with the proposed method, particularly in helicopter and impulsive noise environments.

Figure 4.3 illustrates the comparative results in terms of LSD measures between both methods for various noise types and different input SNRs. From these figures, it can be observed that, whereas the two methods showed comparable improvement in the case of impulsive noise, the estimated LSD values provided by the proposed method were the lowest in all noise conditions.

In terms of segmental SNR, the proposed method can get a performance improvement of about 2 dB on average, over the CSS approach. The largest improvement was achieved in the case of multitalker babble noise, while for impulsive noise this improvement was decreased. This is shown in Figure 4.4.

4.7.2 Speech Spectrograms

Objective measures alone do not provide an adequate evaluation of system performance. Speech spectrograms constitute a well-suited tool for analyzing the time-frequency behavior of any speech enhancement system. All the speech spectrograms presented in this section (Figures 4.5–8) use sentence 1 corrupted with different background noises at SNR = 0 dB.

In the case of white Gaussian noise (Figure 4.5), whereas our method and the CSS approach provide sufficient amount of noise reduction, the spectrum of the former preserves better the desired speech components. In the case of helicopter rotor noise (Figure 4.6), large residual noise components are observed in the spectrograms of the signals processed by the CSS approach. Unlike this method, the spectrogram of the signal processed by our method indicates that the noise between the speech periods is noticeably reduced, while the shape of the speech periods is nearly unchanged. In the case of impulsive noise (Figure 4.7), it can be observed that the CSS approach is less effective for this type of noise. On the contrary, the spectrogram of the signal processed by our method shows that the impulsive noise is moderately reduced in both the speech and noise periods. In the case of multitalker babble noise (Figure 4.8), it can be seen that the CSS approach achieves limited noise reduction, particularly in the noise only periods. By contrast, a good noise reduction was achieved by our method on the entire spectrum.

We can conclude that, while the CSS method achieves limited noise reduction, especially for highly nonstationary noise such as multitalker babble, our method can deal efficiently with both stationary and transient noises with less spectral distortion even in severe noisy environments.

4.7.3 Subjective Listening Tests

In order to validate the objective performance evaluation, subjective listening tests were conducted with our method and the CSS approach. The different considered noise types were added to utterances of the five sentences listed before with SNRs of −5, 0, and 5 dB. The test signals were recorded on a portable computer, and headphones were used during the experiments.

The seven-grade comparison category rating (CCR) was used [32]. The two methods were scored by a panel of twelve subjects asked to rate every sequence of two test signals between −3 and 3. A negative score is given whenever the former test signal sound more pleasant and natural to the listener than the latter. Zero is selected if there is no difference between the two test signals. For each subject, the following procedure was applied: 1) each sequence of two test signals was played with brief pauses in between tracks and repeated twice in a random order; 2) the listener was then asked if he wished to hear the current sequence once more or skip to the next. This led to 60 scores for each test session which took about 25 minutes per subject.

The results, averaged over the 12 listeners' scores and the 5 test sentences, are shown in Figure 4.9. For the considered background noises, CCRs ranging from 0.33 to 1.27

were achieved over the alternative approach. The maximum improvement of CCR was obtained in the case of helicopter noise (1.1) and multitalker babble noise (1.27), while the worst score was achieved for additive white noise (0.33). The reason behind the roughly similar performance of the two methods in the case of white noise can be understood by recognizing that the minimum statistics noise PSD estimator performs better in the presence of stationary noise as opposed to nonstationary noise.

4.8 Conclusion

Given two received signals corrupted by additive noise, using the minimum statistics method to estimate noise after Zelinski's noise reduction approach, can substantially reduce the residual and coherent noise components that would otherwise be present at the output of Zelinski's filter. Objective evaluation results show that a performance improvement in terms of segmental SNR of about 2 dB on average can be achieved over the CSS approach. The best noise reduction was obtained in the case of multitalker babble noise, while the improvement was lower for impulsive noise. Subjective listening tests performed on a limited data set revealed that CCRs ranging from 0.33 to 1.27 can be achieved over the CSS approach. The maximum improvement of CCR was obtained in the case of helicopter and multitalker babble noises, while the worst score was achieved when white noise was added.

A fruitful direction of further research would therefore be to extend the method to multiple microphones as well as to investigate the benefits of such extension on the overall system performance.

ACKNOWLEDGMENTS

The authors wish to thank M. Boukadoum, UQAM, Montreal, for his help in commenting constructively the proposed method. The authors also acknowledge cooperation of the volunteers who took part in the subjective listening tests reported in this paper, in particular, B. Girodias, Ecole Polytechnique de Montreal.

REFERENCES

[1].J. Benesty, S. Makino, and J. Chen, *Speech Enhancement*, Springer, New York, NY, USA, 2005.

[2].Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. on ASSP*, vol. 32, no. 6, pp. 1109-1121, 1984.

[3].I. Cohen, "Multichannel post-filtering in nonstationary noise environments," *IEEE Trans. on SP*, vol. 52, no. 5, pp. 1149–1160, 2004.

[4].J. V. Berghe and J. Wooters, "An adaptive noise canceller for hearing aids using two nearby microphones," *Journal of the Acoustical Society of America*, vol. 103, no. 6, pp. 3621–3626, 1998.

[5].J.-B. Maj *et al.*, "Comparison of adaptive noise reduction algorithms in dual microphone hearing aids," *Speech Communication*, vol. 48, no. 8, pp. 957–970, 2006.

[6]. R. Le Bouquin-Jannès *et al.*, "Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator," *IEEE Trans. on SAP*, vol. 5, pp. 484–487, 1997.

[7]. A. Guerin *et al.*, "A two-sensor noise reduction system: applications for hands-free car kit," in *EURASIP Journal on Applied Signal Processing*, pp. 1125–1134, 2003.

[8]. X. Zhang and Y. Jia, "A soft decision based noise cross power spectral density estimation for two-microphone speech enhancement systems," *IEEE IC on ASSP*, vol. 1, pp. I/813–16, 2005.

[9]. D. O'Shaughnessy, *Speech communications, human and machine*, New York: IEEE Press, 2000.

[10]. Y. Kaneda and M. Tohyama, "Noise suppression signal processing using 2-point received signal," *Electronics and Communications in Japan*, vol. 67–A, pp. 19–28, 1984.

[11]. R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," *in Proc. 13th IEEE Int. Conf. on ASSP*, vol. 5, pp. 2578–2581, 1988.

[12]. ——, "Noise reduction based on microphone array with LMS adaptive post-filtering," *Electronic Letters*, vol. 26, no. 24, pp. 2036–2581, 1990.

[13]. I.A. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Trans. on SAP*, vol. 11, no. 6, pp. 709–716, 2003.

[14]. K.U. Simmer and A. Wasiljeff, "Adaptive microphone arrays for noise suppression in the frequency domain," *in Proc. 2nd COST-229 Worshop on Adaptive Algorithms in Communications*, pp. 185–194, 1992.

[15]. S. Lefkimmiatis and P. Maragos, "A generalized estimation approach for linear and nonlinear microphone array post-filters," *Speech Communication*, vol. 49, pp. 657–666, 2007.

[16]. L.J. Griffiths and C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, 1982.

[17]. S. Fischer and K.U. Simmer, "Beamforming microphone arrays for speech acquisition in noisy environments," *Speech Communication*, vol. 20, no. 3–4, pp. 215–227, 1996.

[18]. J. Bitzer *et al.*, "Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement," *in Proc. 24th IEEE Int. Conf. on ASSP*, vol. 5, pp. 2965–2968, 1999.

[19]. S. Fischer and K.-D. Kammeyer, "Broadband beamforming with adaptive postfiltering for speech acquisition in noisy environments," *in Proc. 22th IEEE Int. Conf. on ASSP*, vol. 1, pp. 359–362, 1997.

[20]. C. Marro *et al.*, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. on SAP*, vol. 6, no. 3, pp. 240–259, 1998.

[21]. I. Cohen *et al.*, "An integrated real-time beamforming and postfiltering system for nonstationary noise environments," in *EURASIP Journal on Applied Signal Processing*, pp. 1064–1073, 2003.

[22]. K. U. Simmer *et al.*, "Suppression of coherent and incoherent noise using a microphone array," *Annales des Télécommunications*, vol. 49, pp. 439–446, 1994.

[23]. R. Martin, "Noise power spectral estimation based on optimal smoothing and minimum statistics," *IEEE Trans. on SAP*, vol. 9, pp. 504–512, 2001.

[24]. I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Trans. on SAP*, vol. 9, no. 1, pp. 12–15, 2002.

[25]. I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Trans. on SAP*, vol. 11, no. 5, pp. 466–475, 2003.

[26]. R. Martin, "Bias compensation methods for minimum statistics noise power spectral density estimation," *Signal Processing*, vol. 86, no. 6, pp. 1215–1229, 2006.

[27]. D. Mauler and R. Martin, "Noise power spectral density estimation on highly correlated data," *In Proc. of the 10th IWAENC*, Sep. 12–14, 2006.

[28]. M. Nilsson et al., "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.*, vol. 95, no. 2, pp. 1085-1099, 1994.

[29]. F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Trans. on ASSP*, vol. 23, pp. 67-72, 1975.

[30]. U. Mittal and N. Phamdo, "Signal/Noise KLT based approach for enhancing speech degraded by colored noise," *IEEE Trans. on SAP*, vol. 8, no. 2, pp. 159-167, 2000.

[31]. S. Quakenbush et al., *Objective Measures of Speech Quality*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[32]. I TU-T, Recommendation P.800, *Methods for subjective determination of transmission quality*. International Telecommunication Union Radiocommunication Assembly, 1996.

Figures

TABLE 4.1 COMPARATIVE PERFORMANCE IN TERMS OF MEAN ITAKURA-SAITO DISTANCE MEASURE FOR FOUR TYPES OF NOISE AND DIFFERENT INPUT SNRS

SNR_{in} [dB]	White Noise			Helicopter Noise			Impulsive Noise			Babble Noise		
	CSS	MZA	Noisy	CSS	MZA	Noisy	CSS	MZA	Noisy	CSS	MZA	Noisy
–8	1.88	0.62	3.29	2.81	1.92	3.28	2.71	2.03	3.23	2.38	1.26	3.1
–4	1.4	0.43	2.82	2.18	1.29	2.62	2.21	1.67	2.65	1.7	0.85	2.62
0	0.78	0.3	2.23	1.72	0.95	2.18	1.7	1.21	2.06	1.28	0.59	2.12
4	0.51	0.24	1.64	1.28	0.71	1.7	1.34	0.93	1.56	0.92	0.46	1.73
8	0.34	0.25	1.18	0.87	0.47	1.24	0.99	0.69	1.09	0.67	0.32	1.27

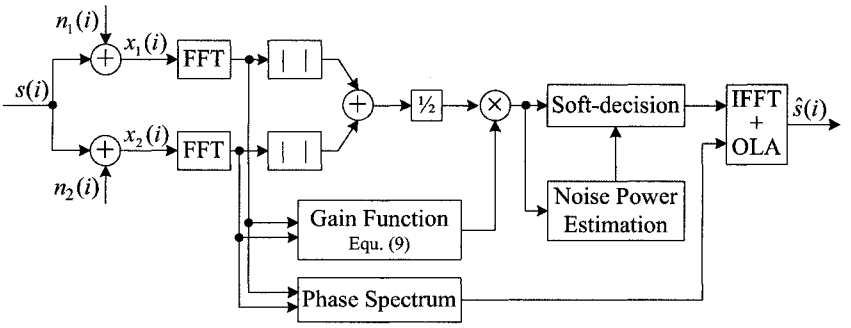

Figure 4.1 The proposed two-microphone algorithm for speech enhancement, where "| |" denotes the magnitude spectrum.

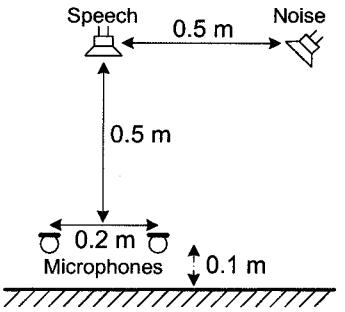

Figure 4.2 Overhead view of the experimental environment.

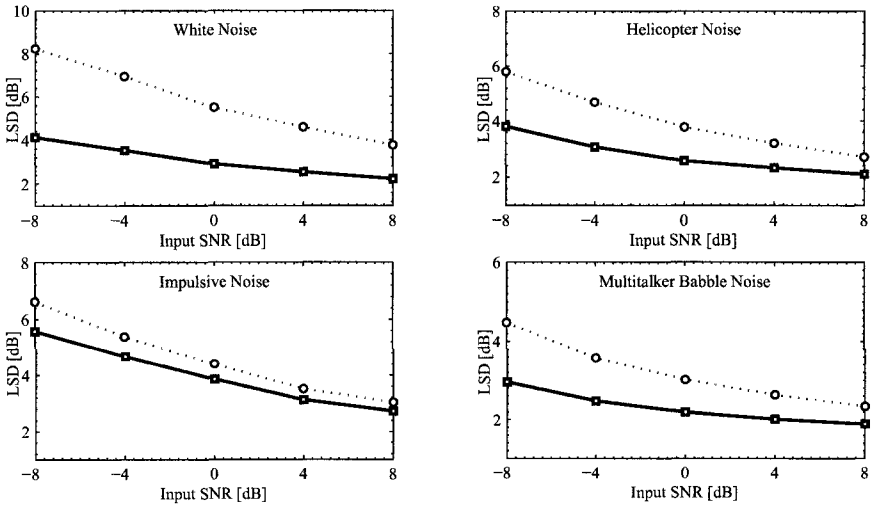

Figure 4.3 Log spectral distortion measure for various noise types and levels, obtained using (○) CSS approach, and (□) the proposed method (MZA).

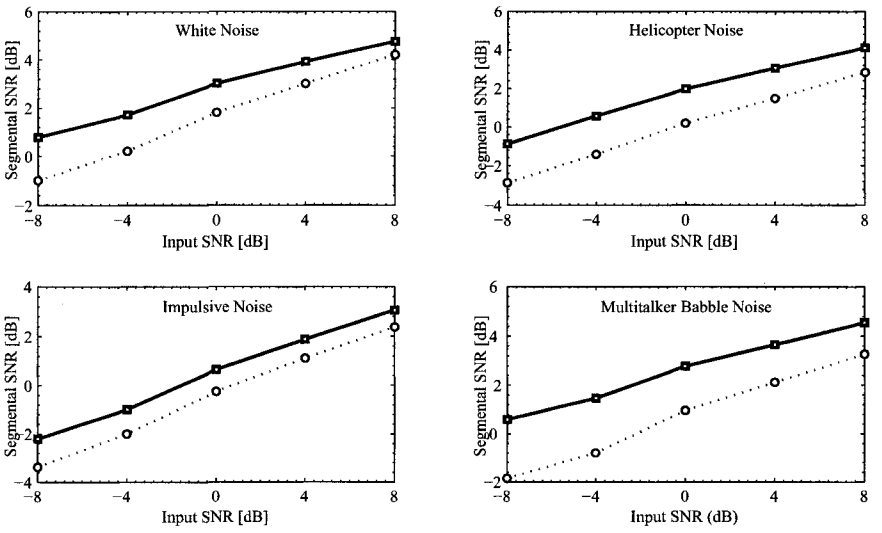

Figure 4.4 Segmental SNR improvement for various noise types and levels, obtained using (○) CSS approach, and (□) the proposed method (MZA).

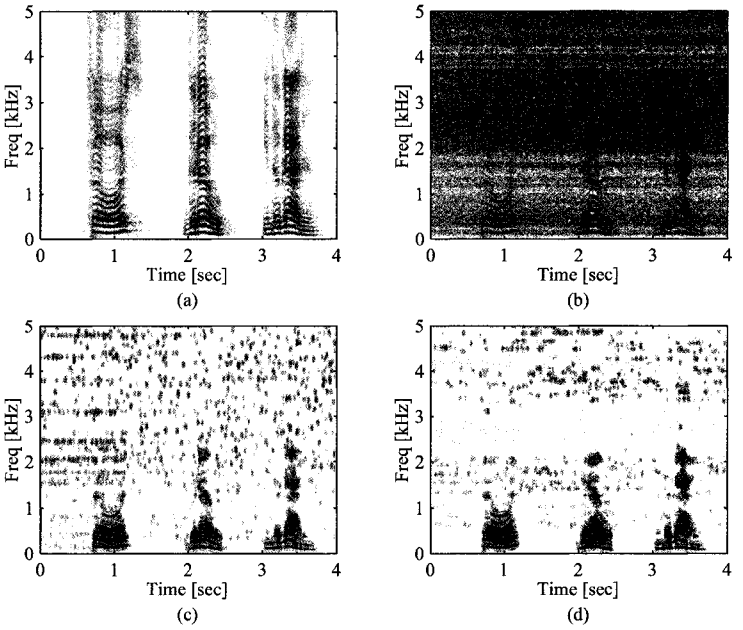

Figure 4.5 Speech spectrograms obtained with white Gaussian noise added at SNR = 0 dB.

(a) Clean speech. (b) Noisy signal. (c) CSS output. (d) MZA output.

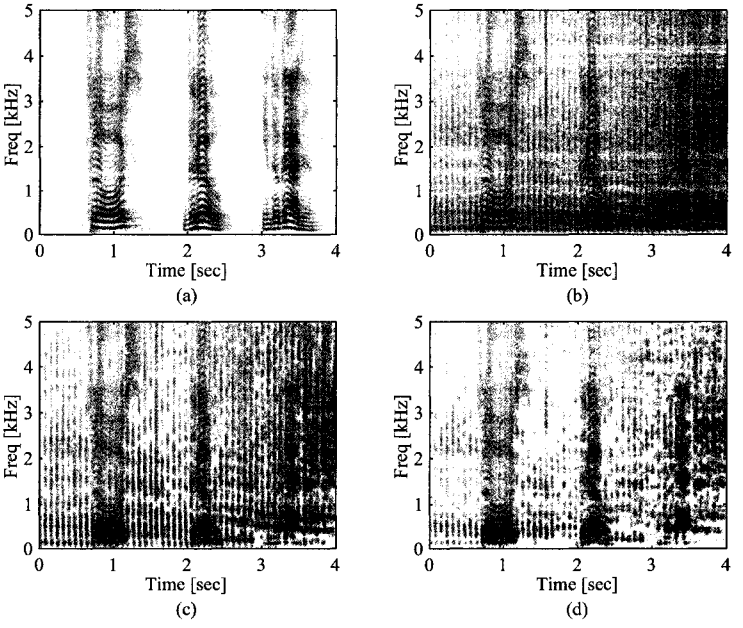

Figure 4.6 Speech spectrograms obtained with helicopter rotor noise added at SNR = 0 dB.

(a) Clean speech. (b) Noisy signal. (c) CSS output. (d) MZA output.

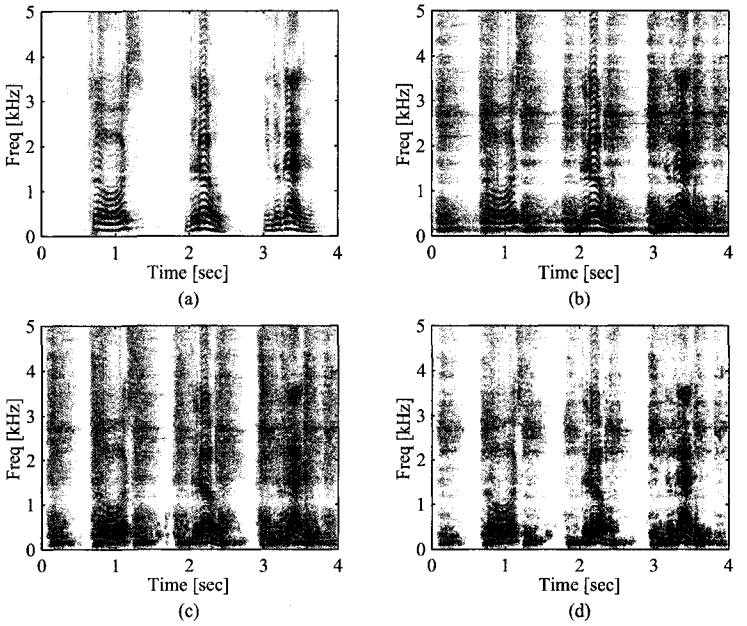

Figure 4.7 Speech spectrograms obtained with impulsive noise added at SNR = 0 dB. (a) Clean speech. (b) Noisy signal. (c) CSS output. (d) MZA output.

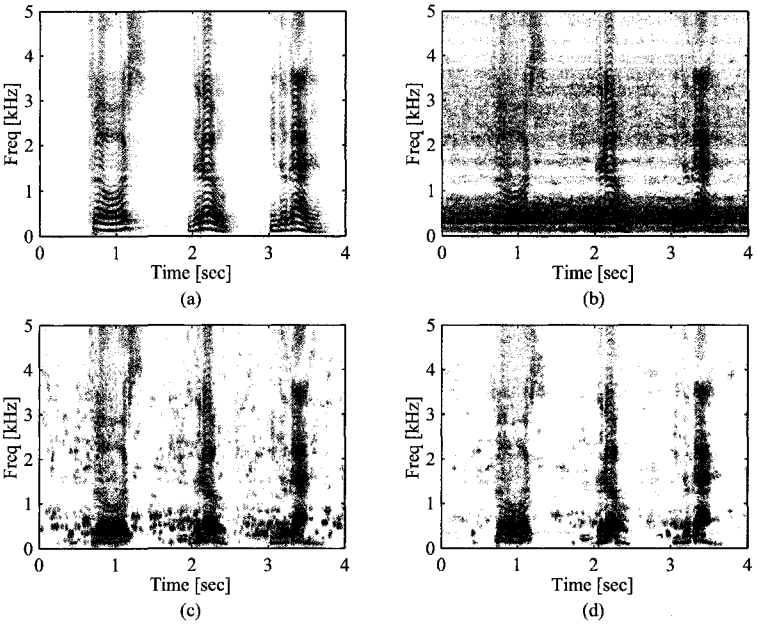

Figure 4.8 Speech spectrograms obtained with multitalker babble noise added at SNR = 0 dB. (a) Clean speech. (b) Noisy signal. (c) CSS output. (d) MZA output.

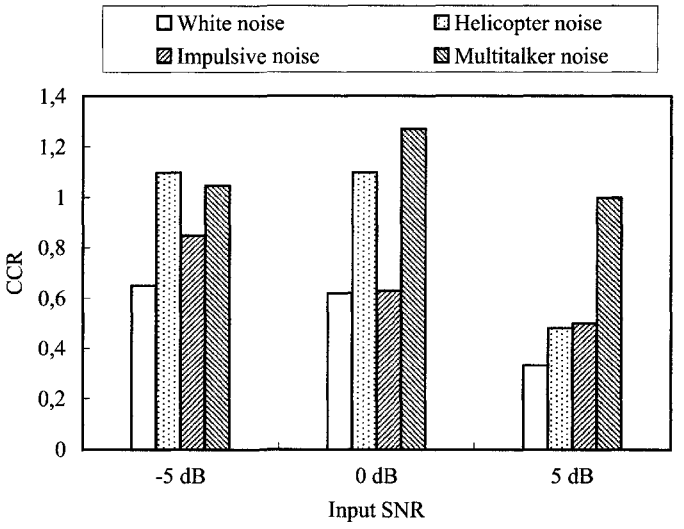

Figure 4.9 CCR improvement against CSS for various noise types and different SNRs.

CHAPITRE 5
RÉDUCTION DE BRUIT DANS LE DOMAINE DE CORRÉLATION

Dans ce chapitre, nous proposons deux versions d'une méthode itérative, élaborée dans le domaine de corrélation, qui permet de limiter la distorsion de l'enveloppe spectrale d'un signal audio large bande corrompu par un bruit de fond additif. Nous discutons à travers ce chapitre les particularités qui différencient les deux versions proposées, lesquelles sont développées dans deux articles qu'on retrouve en annexe. Nous présentons également une méthode d'estimation de la puissance du bruit dans les retards de corrélation considérés. La connaissance de la puissance du bruit est nécessaire pour la compensation des effets de ce dernier sur les paramètres spectraux à évaluer.

5.1 Estimation de la puissance du bruit

Dans la méthode proposée, la puissance du bruit est estimée au moyen d'une version rapide de l'algorithme de MS (pour « Minimum Statistics ») que nous avons implémenté. Rappelons que la méthode de MS est basée sur deux hypothèses essentielles [36]. Premièrement, le signal d'intérêt et celui du bruit sont considérés deux processus aléatoires statistiquement indépendants. Deuxièmement, le niveau d'énergie d'un signal corrompu est fréquemment réduit à des valeurs représentatives du niveau d'énergie du bruit même pendant les périodes d'activité sonore. Cette situation pourra avoir lieu notamment pendant les pauses ou dans de brèves périodes entre les mots et les syllabes.

Ces deux hypothèses rendent donc possible l'obtention d'une estimation de la densité spectrale de puissance (DSP) du bruit. Cette estimation de la DSP du bruit est obtenue en évaluant le minimum, fréquence par fréquence, des DSP dans une fenêtre coulissante contenant des trames de données successives. Cette fenêtre doit être suffisamment large (environ 1 seconde) afin de contenir des pauses ou de brèves périodes d'inactivité sonore. Étant donné que le minimum d'un ensemble de variables aléatoires est plus petit que leur valeur moyenne, la méthode de MS nécessite un facteur de compensation. Ce facteur de compensation est fonction de la variance de la DSP estimée du signal corrompu (bruité). Cette variance est normalisée par la dernière valeur estimée de la DSP du bruit.

La technique développée implémente une version rapide de la méthode de MS. La recherche du minimum est effectuée en fractionnant la fenêtre coulissante de largeur D en deux sections de $D/2$ échantillons de DSP chacune. Pour chaque trame de données, une estimation de la DSP du bruit est obtenue en calculant le minimum entre une valeur actuelle estimée et celle déterminée pendant l'analyse de la section précédente et sauvegardée en mémoire. À la fin de chaque cycle de traitement d'une section de D/2 échantillons de DSP, la valeur de la DSP du bruit en mémoire est actualisée par le minimum des DSP estimées dans cette section.

Il convient de mentionner qu'en implémentant cette version de la méthode de MS, le retard le plus élevé qui pourrait se produire en réponse à une augmentation de la puissance du bruit est approximativement égale à la largeur de la fenêtre coulissante (D).

Il convient également de mentionner que la méthode de MS a tendance à sous-estimer la DSP du bruit, en particulier lorsqu'un taux de chevauchement considérable, entre les trames de données, est utilisé (voir section 4.6.1). Pour remédier à ce problème, nous avons ajusté le facteur de compensation utilisé dans la méthode de MS par une constante décidée empiriquement (voir section 4.6.2).

La puissance du bruit dans les retards de corrélation est ainsi obtenue par la transformée de Fourier inverse du minimum estimé.

5.2 Compensation des effets du bruit

Après avoir estimé la puissance du bruit dans les retards de corrélation, on voudrait pouvoir compenser les effets du bruit sur les paramètres spectraux à évaluer. La plupart des méthodes proposées dans la littérature utilisent les LOYWE pour compenser les effets du bruit sur les paramètres spectraux. Parmi ces méthodes, on retrouve celles qui compensent uniquement le retard de corrélation d'ordre zéro ou supposent que la puissance du bruit est connue. J'ai donc développé une méthode qui repose sur une condition d'arrêt prédéfinie et qui est particulièrement appropriée aux bruits de fond dont les effets s'étendent sur l'ensemble des retards de corrélation (principe illustré à la Figure 5.1). Cette condition d'arrêt n'est satisfaite que lorsque la matrice de corrélation compensée est définie positive.

La méthode proposée consiste à soustraire progressivement de la fonction de corrélation du signal observé (corrompu) une fraction de celle du bruit selon l'équation suivante :

$$\widetilde{R}_{ss}(k) = \left| \, | R_{xx}(k) | -(1-\mu \cdot i) \cdot | R_{xx}(k) | \cdot u(k) \right|$$

$$\cdot \operatorname{sgn}\left\{ | R_{xx}(k) | -(1-\mu \cdot i) \cdot | R_{xx}(k) | \cdot u(k) \right\} \qquad (5.1)$$

où i est un opérateur permettant d'itérer la procédure de compensation, $u(k)$ est la fonction échelon-unité et μ (> 0) est un paramètre représentant le pas de convergence. Une valeur trop faible de ce paramètre permet une compensation plus fine des effets du bruit, mais rend plus lente la convergence de la procédure. Une valeur trop élevée de ce paramètre permet d'accélérer la convergence de la procédure, mais elle risque de produire une réduction insuffisante des effets du bruit. Les tests objectifs que nous avons réalisés sur plusieurs signaux audio contaminés par différents types de bruit pour différents RSB, ont révélé que dans la plupart des cas, un pas de convergence de 0.05 permet un bon compromis entre le taux de convergence et la précision de la compensation. Ceci correspond à un nombre d'itération de 20. Il convient de mentionner que le symbole « sgn » utilisé dans l'équation (5.1) représente la fonction signe qui prend la valeur « +1 » ou « -1 » selon le signe de son argument. Cette fonction permet d'éviter une surestimation des retards de corrélation compensés (en permettant constamment une soustraction entre les pairs de retards de corrélation), particulièrement ceux d'ordres supérieurs où la fonction de corrélation du bruit peut fréquemment prendre des valeurs négatives.

Notre méthode considère l'utilisation des amplitudes des coefficients de réflexion de la matrice de corrélation comme condition d'arrêt prédéfinie. Des paramètres spectraux compensés et stables peuvent ainsi être obtenus aussi longtemps que les coefficients de

réflexion estimés sont strictement inférieurs à l'unité en amplitude. Il convient de mentionner que contrairement à la méthode qui sera présentée dans la prochaine section (section 5.3), les retards de corrélation estimés à partir du signal observé sont compensés des effets d'un bruit de fond à chaque itération.

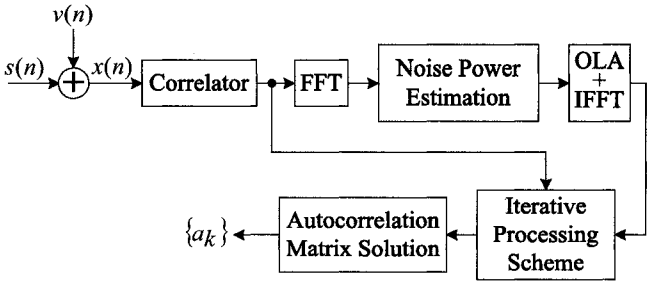

Figure 5.1 Principe de la méthode proposée.

Une structure permettant que l'estimateur de la puissance du bruit et la technique itérative soient réalisés conjointement dans le domaine de corrélation est proposée, permettant ainsi une garantie effective de la stabilité des paramètres spectraux obtenus après réduction de bruit. Cette approche a fait l'objet d'un article de conférence [50]. Une copie de l'article, dans son format original pour *IEEE Northeast Workshop on Circuits and Systems*, pp. 93–96, 5–8 Aug. 2007, se trouve en annexe A.

5.2.1 Contributions principales

Les contributions principales de cette première version de la méthode proposée sont résumées dans les points suivants :

- Élaborée dans le domaine de corrélation, cette méthode permet de garantir la stabilité des paramètres spectraux obtenus après réduction de bruit.

- La méthode proposée est particulièrement appropriée aux bruits de fond dont les effets s'étendent sur l'ensemble des retards de la fonction de corrélation (cas des applications qui utilisent un traitement de préaccentuation avant de procéder à l'analyse LPC d'un signal audio).
- La méthode est avantageuse en temps de calcul et ressources mémoire dans la mesure où elle s'applique sur un nombre relativement faible de retards de corrélation, pour un microphone unique.
- La complexité calculatoire de la méthode est approximativement de $O(p^3)$. Cette complexité calculatoire est nettement inférieure à celle de l'estimation de la fonction de corrélation (qui domine donc l'effort de calcul dans ce type de traitement).

5.2.2 Résultats expérimentaux

La performance de la première version de la méthode proposée a été évaluée sur la phrase « Flowers grow in the garden », prononcée par un locuteur mâle (étudiant au sein de notre département) et échantillonnée à 11.025 kHz. Deux types de bruit, blanc et impulsif, ont été superposés acoustiquement (une légère coloration du bruit blanc est obtenue, selon l'acoustique de l'environnement de l'expérience) à ce signal pour un RSB variant entre –5 dB et +15 dB avec un pas de 5 dB. L'ordre de l'analyse LPC ainsi que la largeur de la fenêtre d'analyse ont été fixés à 15 et environ 23 ms, respectivement. Le facteur de compensation utilisé dans la méthode de MS a été multiplié par la valeur de 2.8 (voir section 5.1). Étant donné que le minimum d'un ensemble de variables aléatoires est plus petit que leur valeur moyenne, un seuil statistique (équation (4.11a)) a

été utilisé pour supprimer les composantes du bruit éventuellement au dessus de la moyenne. Dans toutes nos expériences, ce seuil a été fixé à 5 dB (voir section 4.6.2). Cette valeur a été obtenue empiriquement à travers des tests d'écoute réalisés sur plusieurs signaux audio contaminés par différents types de bruit pour différents RSB. Deux mesures quantitatives ont été considérées : la distance cepstrale et le spectre LPC.

TABLE 5.1 PERFORMANCE EN TERMES DE DISTANCE CEPSTRALE (BRUIT BLANC)

SNR (dB)	**–5**	**0**	**5**	**10**	**15**
Vowel /o/	–1.32	–1.87	–2.15	–1.93	–1.75
Vowel /i/	–0.69	–0.81	–1.05	–1.37	–1.53
Vowel /a/	–1.42	–2.05	–2.31	–2.69	–2.94

TABLE 5.2 PERFORMANCE EN TERMES DE DISTANCE CEPSTRALE (BRUIT IMPULSIF)

SNR (dB)	**–5**	**0**	**5**	**10**	**15**
Vowel /o/	–0.53	–0.84	–1.08	–0.95	–0.79
Vowel /i/	–0.31	–0.43	–0.58	–0.64	–0.73
Vowel /a/	–0.64	–1.02	–1.34	–1.69	–1.88

L'évaluation de la performance de notre méthode, en termes de distance cepstrale, a été effectuée selon la procédure suivante :

- On calcule la distance cepstrale C_{ci} entre le spectre du signal propre et celui traité.
- On calcule la distance cepstrale C_{cd} entre le spectre du signal propre et celui degradé.
- On calcule la différence C_I entre C_{ci} et C_{cd}.

- On considère une amélioration pour des valeurs de $C_I < 0$, et une dégradation de la performance pour des $C_I > 0$.

Mentionnons que dans le calcul des différentes distances cepstrales, le coefficient cepstral d'ordre zéro, qui représente l'énergie moyenne dans la trame de données en analyse, n'a pas été pris en compte. Les tableaux 5.1 et 5.2 résument les résultats obtenus.

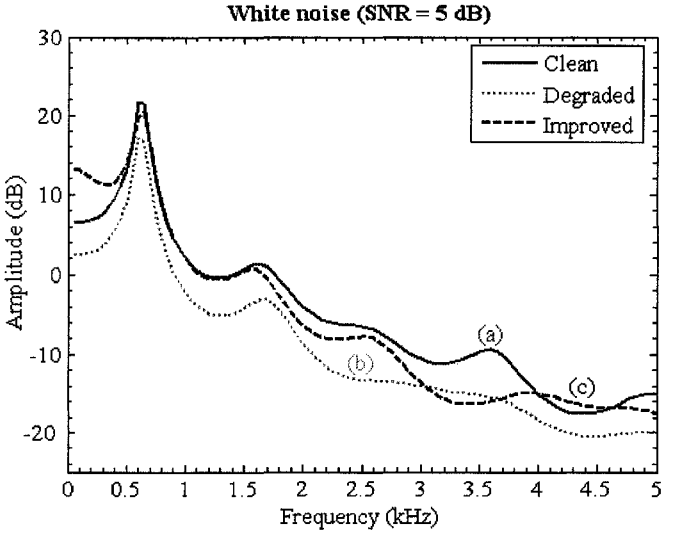

Figure 5.2 Spectres LPC de (a) signal original, (b) signal dégradé, et (c) signal traité, en présence d'un bruit blanc (RSB = 5 dB).

D'après ces résultats, on constate une certaine robustesse au bruit acquise par les paramètres spectraux, par application de la méthode proposée, même pour des conditions bruyantes sévères. Il convient de noter que pour le cas d'un bruit blanc, la performance de la méthode est supérieure à celle pour un bruit impulsif. Ce résultat est

dû notamment à une meilleure estimation, par la méthode de MS, de la puissance d'un bruit stationnaire par opposition à un bruit non stationnaire.

La Figure 5.2 illustre la superposition de différentes estimations d'enveloppes spectrales, notamment par la méthode LPC standard et par la méthode proposée. Cette figure a été obtenue en moyennant 10 réalisations de la seconde occurrence de la voyelle /o/ dans la phrase considérée. Dans ce test, le RSB a été fixé à 5 dB. Ce graphique montre que l'enveloppe spectrale du signal audio traité par notre méthode (courbe (b) sur la Figure 5.2) modélise convenablement la structure formantique du signal original en comparaison avec celle du signal dégradé obtenue par la méthode LPC standard (courbe (c) sur la Figure 5.2). Rappelons que les 3 premiers formants sont particulièrement importants en termes perceptuels dans de nombreuses applications audio (codeurs de voix) [14].

5.3 Amélioration de la procédure de compensation

Bien que la méthode présentée dans la section 5.2 soit particulièrement appropriée à l'ajustement de l'enveloppe spectrale d'un signal audio dégradé par la présence de bruit, elle ne permet en effet qu'une faible réduction de bruit, due à la contrainte d'obtenir, à chaque itération, une matrice de corrélation définie positive pour l'ensemble des retards de corrélation compensés.

Dans la présente section, nous proposons de relaxer cette contrainte en permettant de compenser individuellement les retards de corrélation, en commençant par le retard d'ordre zéro, des effets d'un bruit de fond. Dans cette nouvelle version de la méthode,

lorsqu'un retard de corrélation d'ordre k est en cours de compensation, les retards d'ordres $l < k$ (déjà compensés) et $l > k$ (non encore compensés) sont maintenus inchangés. Ainsi, une meilleure réduction de bruit peut être obtenue sans compromettre la propriété de la matrice de corrélation d'être définie positive. Cette approche a fait l'objet d'un article de conférence [51]. Une copie de l'article, dans son format original pour *IEEE International Conference on Electronics, Circuits and Systems*, pp. 1364–1367, 11–14 Dec. 2007, se trouve en annexe B.

5.3.1 Contributions principales

Les contributions principales de cette deuxième version de la méthode proposée sont résumées dans les points suivants :

- Élaborée dans le domaine de corrélation, cette méthode permet une meilleure réduction de bruit (en comparaison avec celle présentée dans la section 5.2) tout en garantissant la stabilité des paramètres spectraux.
- La méthode est avantageuse en temps de calcul et ressources mémoire dans la mesure où elle s'applique sur un nombre relativement faible de retards de corrélation, pour un microphone unique.
- L'effort calculatoire requit par la méthode est approximativement de $O(p^4)$. Pourtant significatif, cet effort calculatoire ne devrait pas poser de problèmes pour les signaux audio compte tenu de leur largeur de bande relativement étroite (p étant sélectionné selon l'équation (2.24)) et l'utilisation croissante d'architectures de traitement parallèles dans la plupart des applications de traitement du signal.

5.3.2 Résultats expérimentaux

La performance de la deuxième version de la méthode proposée a été évaluée selon le même protocole expérimental que celui de la section 5.2.2. Rappelons que nous avons considéré l'utilisation de la phrase « Flowers grow in the garden », prononcée par un locuteur mâle. Deux types de bruit, blanc et impulsif, ont été superposés acoustiquement (une légère coloration du bruit blanc est obtenue, selon l'acoustique de l'environnement de l'expérience) à ce signal pour un RSB variant entre –5 dB et +15 dB avec un pas de 5 dB. L'ordre de l'analyse LPC ainsi que la largeur de la fenêtre d'analyse ont été fixés à 15 et environ 23 ms, respectivement. Deux mesures quantitatives ont été considérées : la distance cepstrale et le spectre LPC.

La Figure 5.3 illustre la superposition de différentes estimations d'enveloppes spectrales, notamment par la méthode LPC standard et par la méthode proposée. Cette figure a été obtenue en moyennant 10 réalisations de la seconde occurrence de la voyelle /o/ dans la phrase considérée. Dans ce test, Un bruit blanc est superposé au signal original pour un RSB de 0 dB. Ce graphique montre que l'enveloppe spectrale du signal audio traité par notre méthode (courbe (b) sur la Figure 5.2) modélise convenablement la structure formantique du signal original en comparaison avec celle du signal dégradé obtenue par la méthode LPC standard (courbe (c) sur la Figure 5.2). Rappelons que les 3 premiers formants sont particulièrement importants en termes perceptuels dans de nombreuses applications audio (codeurs de voix) [14]. Des résultats similaires ont été obtenus avec un bruit impulsif superposé au signal original pour un RSB de 0 dB (Figure 5.4).

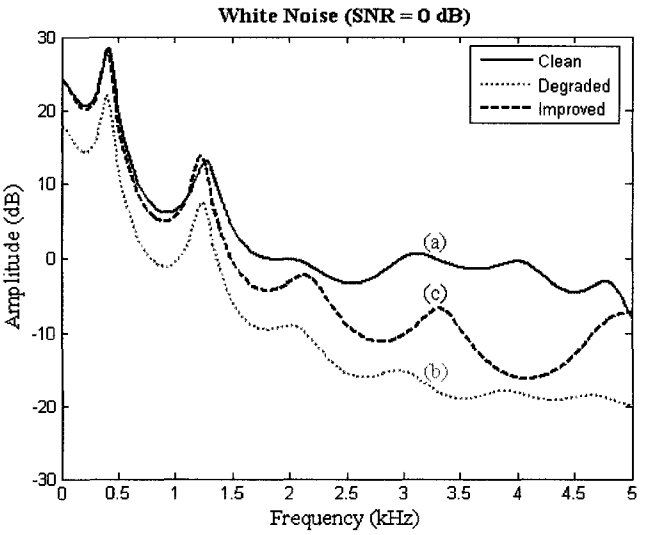

Figure 5.3 Spectres LPC de (a) signal original, (b) signal dégradé, et (c) signal traité, en présence d'un bruit blanc (RSB = 0 dB).

Toutefois, une accentuation des formants d'ordres supérieurs est observée sur le spectre LPC de la Figure 5.4. Contrairement au bruit blanc gaussien, le bruit impulsif possède une statistique non standard. Ce type de bruit est généralement non stationnaire, non-gaussien et possède un contenu fréquentiel très complexe. Son spectre peut avoir des antirésonances prédominantes aussi bien en basses fréquences qu'en hautes fréquences. Le phénomène observé sur la Figure 5.4 peut donc être attribué à deux facteurs probables : le comportement en hautes fréquences des perturbations impulsives utilisées dans cette expérience (antirésonances) et l'inaptitude de l'algorithme de MS à estimer, avec suffisamment de précision, la puissance d'un bruit non stationnaire par opposition à un bruit stationnaire. Il est intéressant de constater que pour un RSB de 0 dB, la deuxième version de la méthode proposée fournit un résultat similaire à celui obtenu par la première version pour un RSB de 5 dB.

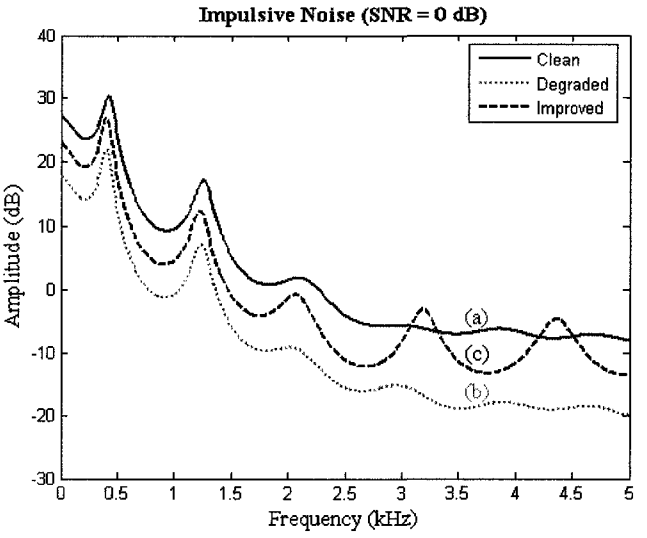

Figure 5.4 Spectres LPC de (a) signal original, (b) signal dégradé, et (c) signal traité, en présence d'un bruit impulsif (RSB = 0 dB).

L'évaluation de la performance de notre méthode, en termes de distance cepstrale, a été effectuée selon la procédure décrite dans la section 5.2.2. Mentionnons que dans le calcul des différentes distances cepstrales, le coefficient cepstral d'ordre zéro, qui représente l'énergie moyenne dans la trame de données en analyse, n'a pas été pris en compte. Le tableau 5.3 résume les résultats obtenus.

D'après ces résultats, on constate une certaine robustesse au bruit acquise par les paramètres spectraux, par application de la méthode proposée, même pour des conditions bruyantes sévères.

TABLE 5.3 PERFORMANCE EN TERMES DE DISTANCE CEPSTRALE (BRUIT BLANC)

SNR (dB)	**–5**	**0**	**5**	**10**	**15**
Vowel /o/	–1.37	–1.91	–2.19	–1.97	–1.81
Vowel /i/	–0.73	–0.85	–1.09	–1.41	–1.55
Vowel /a/	–1.45	–2.08	–2.35	–2.72	–2.98

Il convient de noter que la performance de la deuxième version de la méthode proposée est supérieure à celle de la première version. Ce résultat est dû notamment à la relaxation de la contrainte d'obtenir, à chaque itération, une matrice de corrélation définie positive pour l'ensemble des retards de corrélation compensés (principe de la première version).

CHAPITRE 6
PERSPECTIVES DE DÉVELOPPEMENT

Ce chapitre est dédié aux points qui ont été étudiés, mais qui n'ont fait l'objet d'aucune publication. Quelques perspectives de développement ainsi que des résultats préliminaires et intéressants sont inclus dans ce chapitre, ouvrant la voie à davantage de recherche sur le sujet.

6.1 Traitement paramétrique en présence de bruit

Particulièrement adaptée à la modélisation d'enveloppe spectrale d'un signal audio, l'analyse LPC est retenue dans ce projet de thèse comme méthode paramétrique pour la représentation de l'énergie du spectre d'un signal audio.

Rappelons que l'analyse LPC consiste à extraire de chaque trame de données différents paramètres qui peuvent être regroupés en deux classes distinctes. La première classe regroupe les paramètres spectraux (ou les coefficients PARCOR, pour « partial correlation ») qui représentent l'énergie du spectre d'un signal audio. La seconde classe incorpore un signal résiduel (obtenu par filtrage inverse LPC) qui est souvent associé à un bruit blanc (densité spectrale de puissante égale pour toutes les fréquences).

Le problème de la sensibilité au bruit de fond des traitements paramétriques est bien connu. De faibles variations entre deux trames de données consécutives peuvent entraîner une déviation importante des paramètres spectraux lors de l'analyse. Ces variations sont habituellement générées soit par un bruit ambiant, soit par l'erreur de

quantification. À la synthèse, cette déviation entraîne de fortes variations dans le spectre restitué par le filtre de reconstruction. Ce phénomène d'instabilité conduit souvent à une dégradation globale de la qualité de perception du signal audio reconstruit. Sambur et Jayant [52] ont montré que les dégradations générées par un bruit blanc sont les plus redoutées par le processus de synthèse par LPC.

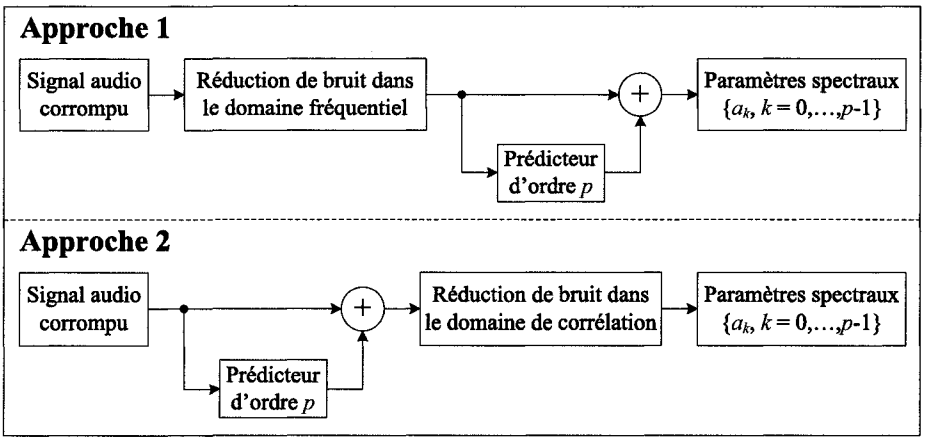

Figure 6.1 Principes des approches de traitement paramétrique en présence de bruit.

Pour pallier ce problème, deux approches de traitement paramétrique en présence de bruit sont souvent utilisées :

1. La première approche consiste à effectuer un traitement pour la réduction de bruit, généralement dans le domaine fréquentiel, comme un prétraitement pour la modélisation paramétrique du signal audio.
2. La seconde approche consiste plutôt à appliquer un traitement pour la réduction de bruit directement dans le domaine de corrélation afin d'évaluer les paramètres spectraux qui modélisent convenablement l'enveloppe spectrale du signal.

Les principes de ces deux approches sont illustrés à la Figure 6.1. Rappelons qu'un des objectifs de cette thèse est de déterminer laquelle des deux approches permet d'obtenir la meilleure fidélité de l'enveloppe spectrale, d'un signal audio, à sa structure formantique.

6.2 Disposition du traitement combiné

Dans cette étude comparative, nous considérons un signal audio contaminé par un bruit blanc gaussien additif. Nous utilisons la méthode développée dans le chapitre 4 (traitement pour la réduction de bruit dans le domaine fréquentiel) comme un prétraitement pour la modélisation paramétrique du signal audio (principe de la première approche illustrée à la Figure 6.1). Quant à la seconde approche (illustrée à la Figure 6.1), elle est implémentée moyennant l'utilisation de l'une des deux méthodes développées dans le chapitre 5 (traitement pour la réduction de bruit dans le domaine de corrélation). Il est intéressant de constater que ces deux méthodes (objets du chapitre 5) sont identiques en cas d'un bruit blanc gaussien additif. Rappelons qu'en cette situation de bruit, seul le retard de corrélation d'ordre zéro est compensé, alors que les retards de corrélation d'ordres supérieurs sont maintenus inchangés.

La Figure 6.2 illustre la superposition du spectre d'un signal de parole (obtenu par FFT) et différentes estimations d'enveloppes spectrales, notamment par la méthode LPC standard et par les approches de traitement combiné considérées. Cette figure est obtenue en moyennant 10 réalisations de la première occurrence de la voyelle /e/ dans la phrase « Flowers grow in the garden », prononcée par un locuteur mâle et échantillonnée

à 11.025 kHz. Un bruit blanc est superposé acoustiquement (une légère coloration du bruit est obtenue, selon l'acoustique de l'environnement de l'expérience) à ce signal pour un RSB de 0 dB. L'ordre de l'analyse LPC ainsi que la largeur de la fenêtre d'analyse sont fixés à 15 et approximativement 23 ms, respectivement. Le facteur de compensation utilisé dans la méthode de MS ainsi que le seuil statistique utilisé dans l'équation (4.11a) ont été fixés comme indiqué dans la section 5.6.2.

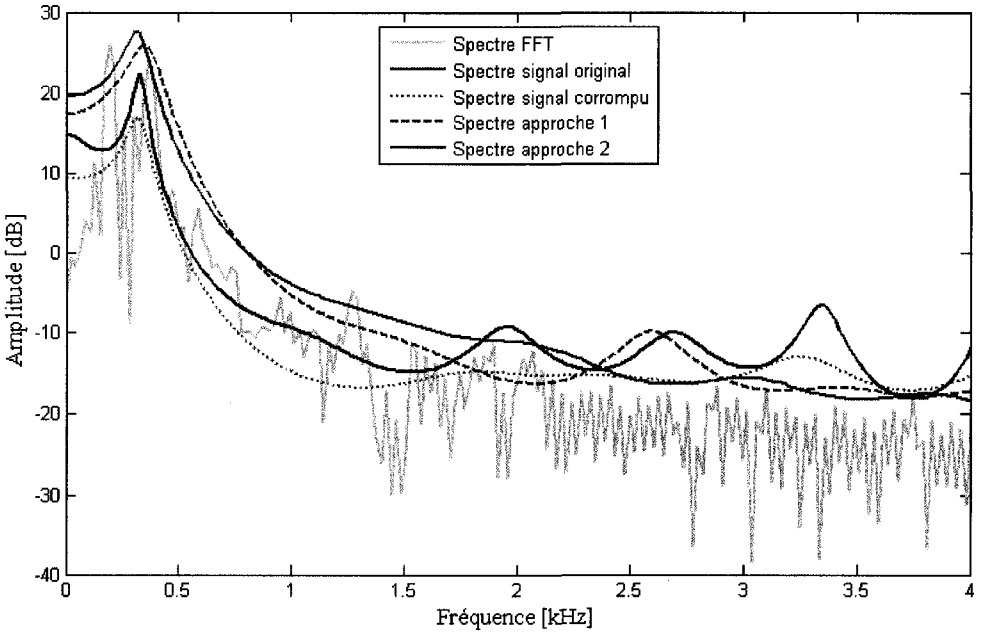

Figure 6.2 Comparaison des approches de traitement combiné.

En comparaison avec la seconde approche, ce graphique montre que l'utilisation d'un traitement pour la réduction de bruit, dans le domaine fréquentiel, comme un prétraitement pour la modélisation paramétrique d'un signal audio (principe de la première approche) permet d'obtenir une meilleure fidélité de l'enveloppe spectrale, d'un signal audio, à sa structure formantique. Ce résultat est cohérent avec les travaux

proposés dans la littérature [53] dans la mesure où la première approche réussit efficacement à réduire le bruit de fond tout en contournant les problèmes d'instabilité souvent rencontrés par les techniques de modélisation linéaire prédictive. Par contre, la seconde approche ne permet qu'une réduction partielle du bruit décidée par la contrainte de stabilité des paramètres spectraux à évaluer.

Il convient de mentionner que nous avons obtenu des résultats similaires sur d'autres phrases et phonèmes. Ces comparaisons étant subjectives sur l'allure d'un grand nombre de graphiques, nous avons présenté dans ce chapitre seulement un exemple.

6.3 Amélioration de la précision de l'estimateur de la variance du bruit

Les méthodes développées dans le chapitre 5 reposent sur l'algorithme de MS [36] pour évaluer la puissance du bruit dans la séquence d'autocorrélation du signal observé. Ces deux méthodes sont particulièrement appropriées à être utilisées dans des situations de bruit dont les effets s'étendent sur l'ensemble des retards de la fonction de corrélation. Deux problèmes bien connus sont associés à l'utilisation de l'algorithme de MS. Premièrement, la variance de la DSP de bruit estimée par cette méthode est deux fois plus élevée que celle d'un estimateur classique [38]. Deuxièmement, en cas de durées suffisamment prolongées des périodes de silence (imposant un débit audio plus lent), des segments audio de faible énergie (c.-à-d., son non voisé) peuvent occasionnellement être confondus avec des niveaux de bruit si la largeur sélectionnée de la fenêtre de recherche du minimum n'est pas convenablement choisie (c.-à-d., n'est pas

assez large). Ainsi, les paramètres spectraux obtenus après compensation peuvent ne pas représenter la structure formantique du signal avec suffisamment de précision.

Pour pallier ces problèmes, nous avons développé une nouvelle méthode qui permet d'améliorer la précision de l'estimateur de la variance du bruit dans le cas d'un signal audio contaminé par un bruit blanc additif. Cette méthode consiste à substituer l'estimateur de la DSP de bruit basé sur l'algorithme de MS par un nouvel estimateur qui repose sur la méthode ODNE de Cadzow [1], la décomposition en valeurs singulières (SVD pour « Singular Value Decomposition ») tronquée de la matrice de corrélation (au sens des moindres carrés) [56] et la propriété statistique d'appariement des retards de corrélation d'ordres supérieurs [55]. Contrairement aux méthodes développées dans le chapitre 5, cette méthode exploite le principe que dans le domaine de corrélation seul le retard de corrélation d'ordre zéro est affecté par un bruit blanc additif, alors que les retards de corrélation d'ordres supérieurs sont maintenus inchangés [42]. Il convient de mentionner que cet estimateur atteint la borne de Cramér-Rao (borne inférieure sur la variance d'un estimateur sans biais) même pour des valeurs de RSB inférieures à 0 dB [55].

Ensuite, nous avons développé une technique itérative qui consiste à réduire par soustraction le bruit affectant le retard de corrélation d'ordre zéro. En plus de l'amplitude des coefficients de réflexion, cette technique considère l'utilisation de la valeur propre minimale de la matrice de corrélation comme condition d'arrêt prédéfinie.

Étant identiques dans le cas d'un bruit blanc gaussien additif, nous comparons l'une des deux méthodes développées dans le chapitre 5 (désignée dans ce chapitre par AMS) à celle qui repose sur la SVD tronquée et la propriété statistique d'appariement des retards de corrélation d'ordres supérieurs (désignée dans ce chapitre par ASVD).

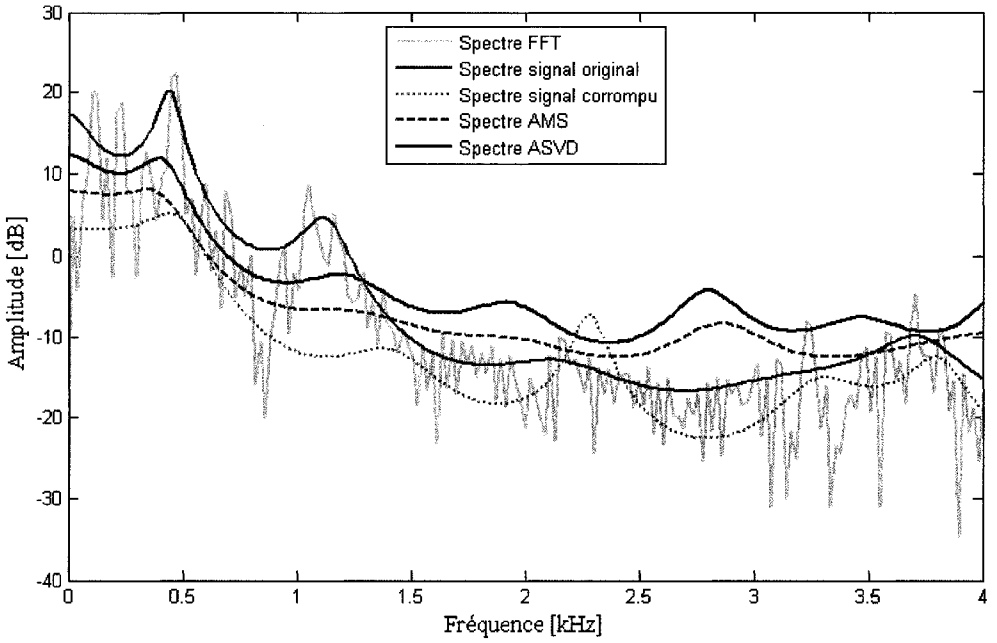

Figure 6.3 Comparaison des méthodes AMS et ASVD.

La Figure 6.3 illustre la superposition du spectre d'un signal de parole (obtenu par FFT) et différentes estimations d'enveloppes spectrales, notamment par la méthode LPC standard, la méthode AMS et la méthode ASVD. Cette figure est obtenue en moyennant 10 réalisations de la première occurrence de la voyelle /o/ dans la phrase « A boy ran down the path », issue de la base de données HINT, prononcée par un locuteur mâle et échantillonnée à 11.025 kHz. Un bruit blanc gaussien est synthétisé par un générateur de bruit et superposé acoustiquement (une légère coloration du bruit est obtenue, selon

l'acoustique de l'environnement de l'expérience) à ce signal pour un RSB de 0 dB. La largeur de la fenêtre d'analyse est fixée à approximativement 23 ms. L'ordre de l'analyse LPC est établi à 15.

Ce graphique montre que l'enveloppe spectrale du signal audio traité par la méthode ASVD (dessinée en ligne continue noire sur la Figure 6.3) représente plus fidèlement l'énergie du spectre (particulièrement en basse fréquence) en comparaison avec celle obtenue par la méthode AMS (dessinée en trait pointillé vert sur la Figure 6.3). Ce résultat est dû notamment à une estimation plus précise de la variance du bruit par la méthode ASVD que celle obtenue par la méthode AMS. Bien évidemment, une estimation précise de la variance du bruit conduit à l'obtention de paramètres spectraux robustes et qui modélisent convenablement la structure formantique du signal.

Les principales contributions de la méthode ASVD proposée, semblent être :

- En comparaison d'un estimateur basé sur l'algorithme de MS, cette méthode permet d'obtenir une estimation plus précise de la variance d'un bruit de fond additif, en particulier blanc.
- Particulièrement appropriée à un bruit blanc centré et à distribution gaussienne, cette méthode fournit des paramètres spectraux stables et précis.
- L'effort calculatoire de la méthode est approximativement de $O(p^3)$ (excluant celui nécessaire à la détermination du rang d'une matrice $p \times q$ au moyen de la SVD

tronquée). Considéré réduit, cet effort calculatoire est d'autant plus faible que le nombre de périodes de silence dans un signal audio est élevé.

Ces premiers résultats, même s'ils restent préliminaires, permettent de maintenir ouvertes des voies de recherche sur le sujet relativement peu exploré par les chercheurs.

CHAPITRE 7
DISCUSSION GÉNÉRALE ET CONCLUSION

En s'articulant autour de trois axes, les travaux de recherche couverts par cette thèse proposent des solutions algorithmiques innovantes et pertinentes pour améliorer la qualité de l'information audio large bande dans des applications portables munis d'un réseau de communication. Nous considérons le cas des aides auditives numériques.

Le premier axe porte sur le développement d'une nouvelle approche à deux microphones qui repose sur la combinaison d'une technique de filtrage adaptatif proposée auparavant dans la littérature et d'un estimateur de DSP du bruit. Élaborée dans le domaine fréquentiel, cette méthode est considérée avantageuse en temps de calcul et ressources mémoire, ce qui permet d'envisager son implémentation en temps réel. Pour évaluer la méthode proposée en termes de performance, nous avons réalisé de nombreux tests objectifs et subjectifs sur plusieurs signaux audio contaminés par différents types de bruit pour différents RSB. Les résultats de ces tests ont démontré la supériorité de notre méthode par rapport à d'autres méthodes concurrentes, notamment dans des situations de bruit fortement non stationnaire (bruit impulsif et bruit d'ambiance de type « cocktail-party »).

Bien que les méthodes à deux microphones soient simples, fiables en termes de consommation d'énergie et faciles à implémenter en temps réel, elles n'offrent des conditions d'écoute optimales pour les malentendants que lorsque le locuteur et le bruit sont diamétralement opposés dans un espace peu réverbérant (cas des microphones

directionnels). De nombreuses études montrent que l'utilisation d'un réseau de microphones permettant de focaliser l'antenne formée vers le locuteur avec qui le malentendant converse, constitue une solution prometteuse pour augmenter la discrimination des sons et améliorer la compréhension de la parole dans le bruit [54]. En ce sens, une voie encourageante serait de généraliser la méthode développée dans le chapitre 4 à un réseau de microphones et procéder à des études comparatives supplémentaires pour confirmer son efficacité en termes de gain en intelligibilité et en agrément d'écoute.

Le second axe concerne l'élaboration d'une approche innovante qui permet de limiter la distorsion de l'enveloppe spectrale d'un signal audio large bande corrompu par un bruit de fond additif. Il s'agit d'une méthode itérative qui repose sur une condition d'arrêt prédéfinie et l'algorithme de MS (pour « Minimum Statistics »). Des paramètres spectraux compensés et stables peuvent ainsi être obtenus aussi longtemps que les coefficients de réflexion estimés sont strictement inférieurs à l'unité en amplitude (la matrice de corrélation compensée est définie positive). Deux versions de cette méthode ont été proposées. La première version ne permet en effet qu'une faible réduction de bruit, due à la contrainte d'obtenir, à chaque itération, une matrice de corrélation définie positive pour l'ensemble des retards de corrélation compensés. Afin de relaxer cette contrainte, nous avons proposé une seconde version de cette méthode qui permet de compenser individuellement les retards de corrélation (en commençant par le retard de corrélation d'ordre zéro) des effets d'un bruit de fond. Ainsi, une meilleure réduction de

bruit peut être obtenue sans compromettre la propriété de la matrice de corrélation d'être définie positive.

En effet, deux problèmes bien connus sont associés à l'utilisation de l'algorithme de MS. Premièrement, la variance de la DSP de bruit estimée par cette méthode est deux fois plus élevée que celle d'un estimateur classique. Deuxièmement, en cas d'un débit audio plus lent, des segments audio de faible énergie peuvent occasionnellement être confondus avec des niveaux de bruit si la largeur sélectionnée de la fenêtre de recherche du minimum n'est pas convenablement choisie (c.-à-d., n'est pas assez large). Ainsi, les paramètres spectraux obtenus après compensation peuvent ne pas représenter la structure formantique du signal avec suffisamment de précision.

Pour pallier ces problèmes, nous avons développé une nouvelle méthode qui permet d'améliorer la précision de l'estimateur de la variance du bruit dans le cas d'un signal audio contaminé par un bruit blanc additif. Cette méthode consiste à substituer l'estimateur de la DSP de bruit basé sur l'algorithme de MS par un nouvel estimateur qui repose sur la méthode ODNE de Cadzow, la décomposition en valeurs singulières (SVD pour « Singular Value Decomposition ») tronquée de la matrice de corrélation (au sens des moindres carrés) et la propriété statistique d'appariement des retards de corrélation d'ordres supérieurs.

Contrairement aux deux méthodes développées dans le chapitre 5, cette méthode exploite le principe que dans le domaine de corrélation seul le retard de corrélation d'ordre zéro est affecté par un bruit blanc additif, alors que les retards de corrélation

d'ordres supérieurs sont maintenus inchangés. Il convient de mentionner que cet estimateur atteint la borne de Cramér-Rao même pour des valeurs de RSB inférieures à 0 dB. Ensuite, nous avons développé une technique itérative qui consiste à réduire par soustraction le bruit affectant le retard de corrélation d'ordre zéro. En plus de l'amplitude des coefficients de réflexion, cette technique considère l'utilisation de la valeur propre minimale de la matrice de corrélation comme condition d'arrêt prédéfinie.

En comparaison d'un estimateur basé sur l'algorithme de MS, cette méthode permet d'obtenir une estimation plus précise de la variance d'un bruit de fond additif, en particulier blanc. Bien évidemment, une estimation précise de la variance du bruit conduit à l'obtention de paramètres spectraux robustes et qui modélisent convenablement la structure formantique du signal. Ces premiers résultats, même s'ils restent préliminaires, permettent de maintenir ouvertes des voies de recherche sur le sujet relativement peu exploré par les chercheurs.

Quant au troisième axe, il porte sur une étude empirique permettant de déterminer l'ordre qui conduit à l'obtention de la meilleure fidélité de l'enveloppe spectrale, d'un signal audio, à sa structure formantique, lorsque les deux traitements, paramétrique et réduction de bruit, sont combinés. Les résultats obtenus sont cohérents avec les travaux proposés dans la littérature dans la mesure où la stratégie de faire précéder le traitement paramétrique classique par un traitement pour la réduction de bruit permet de réduire considérablement le bruit de fond tout en contournant les problèmes d'instabilité souvent rencontrés par les techniques de modélisation linéaire prédictive.

De nombreuses perspectives intéressantes de développement sont possibles à partir de ces idées. Une perspective encourageante serait d'étendre l'estimateur de la variance du bruit décrit dans la section 6.2 à des situations plus représentatives de la réalité, en permettant l'évaluation de la puissance du bruit non seulement dans le retard de corrélation d'ordre zéro, mais aussi dans les retards de corrélation d'ordres supérieurs. Ainsi, la réduction de bruit par soustraction serait nettement plus importante et la méthode serait adéquate pour une large variante de perturbations.

Une autre perspective de développement possible serait de mettre en place une plateforme de test basée sur un codeur LPC (composé d'un analyseur, un codeur, un décodeur et un synthétiseur) et un prototype de prothèse auditive, équipé d'un réseau local, afin d'évaluer les méthodes proposées dans des conditions acoustiques réelles et selon des contraintes de fonctionnement en temps réel.

BIBLIOGRAPHIE

[1]. J. A. Cadzow, "Spectral estimation: an overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, pp. 907–939, 1982.

[2]. M. H. Hayes, Statistical digital signal processing and modeling. John Wiley & Sons, Inc. 1996.

[3]. S. M. Kay, Modern spectrum estimation. Prentice-Hall: Englewood Cliffs, NJ. 1988.

[4]. H. Wold, A study in the analysis of stationary time series. Uppsala, Sweden: Almqvist and Wiksell, 2nd ed. 1954.

[5]. S. Kay, "Spectrum analysis–a modern perspective," Proc. IEEE, vol. 69, pp. 1380–1419, 1981.

[6]. L. R. Rabiner et R. W. Schafer, Digital processing of speech signals. Englewood Cliffs, NJ. Prentice-Hall, 1979.

[7]. N. Levinson, "The Wiener (root mean square) error criterion in filter design and prediction", J. Math. Phys., vol. 25, pp. 261–278, 1947.

[8]. S. M. Kay, "Noise compensation for autoregressive spectral estimates," IEEE Trans. on ASSP, vol. 28, no. 3, pp. 292–303, 1980.

[9]. M Morf et al., "Efficient solution of covariance equations for linear prediction", IEEE Trans. on ASSP, vol. 25, no. 5, pp. 429–433, 1977.

[10]. A. Arcese, "On the method of maximum entropy spectrum estimation", IEEE Trans. on Inform. Theory, vol. 29, no. 1, pp. 161–164, 1983.

[11]. D. N. Swingler, "Frequency errors in MEM processing", IEEE Trans. on ASSP, vol. 28, no 2, pp. 257–259, 1980.

[12]. D. M. Wilkes et J. A. Cadzow, "The effects of phase on high-resolution frequency estimators", IEEE Trans. on SP, vol. 41, no 3, pp. 1319–1330, 1993.

[13]. P. M. Djuric, "A model selection rule for sinusoids in white Gaussian noise", IEEE Trans. on ASSP, vol. 44, no 7, pp. 1744-1751, 1996.

[14]. D. O'Shaughnessy, Speech Communications: Human and Machine, 2nd ed. New York: IEEE Press, 2000.

[15]. J. S. Lim, Speech Enhancement. Prentice-Hall: Englewood Cliffs, NJ. 1983.

[16]. S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. on ASSP, vol. 27, no. 2, pp. 113-120, 1979.

[17]. .P. Lockwood et J. Boudy, "Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and the projection, for robust speech recognition in cars", Speech Comm., vol. 11, no. 2-3, pp. 215-228, 1992.

[18]. J.S. Lim et A.V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," Proc. IEEE, vol. 67, no. 12, pp. 1586-1604, 1979.

[19]. L Y. Kaneda et M. Tohyama, "Noise suppression signal processing using 2-point received signal," Electronics and Communications in Japan, vol. 67–A, pp. 19–28, 1984.

[20]. R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in Proc. 13th IEEE Int. Conf. on ASSP, vol. 5, pp. 2578–2581, 1988.

[21]. ——, "Noise reduction based on microphone array with LMS adaptive post-filtering," Electronic Letters, vol. 26, no. 24, pp. 2036–2581, 1990.

[22]. K. U. Simmer et al., "Suppression of coherent and incoherent noise using a microphone array," Annales des Télécommunications, vol. 49, pp. 439–446, 1994.

[23]. R. Le Bouquin-Jeannes, R. Azirani et A. A. Faucon, "Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator," IEEE Trans. on SAP, vol. 5, pp. 484–487, 1997.

[24]. K.U. Simmer et A. Wasiljeff, "Adaptive microphone arrays for noise suppression in the frequency domain," in Proc. 2nd COST-229 Worshop on Adaptive Algorithms in Communications, pp. 185–194, 1992.

[25]. I.A. McCowan et H. Bourlard, "Microphone array post-filter based on noise field coherence," IEEE Trans. on SAP, vol. 11, no. 6, pp. 709–716, 2003.

[26]. S. Lefkimmiatis et P. Maragos, "A generalized estimation approach for linear and nonlinear microphone array post-filters," Speech Communication, vol. 49, pp. 657–666, 2007.

[27]. L.J. Griffiths et C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. on Antennas and Propagation, vol. 30, no. 1, pp. 27–34, 1982.

[28]. A. Guerin, R. Le Bouquin-Jeannes et A. A. Faucon, "A two-sensor noise reduction system: applications for hands-free car kit," in EURASIP Journal on Applied Signal Processing, pp. 1125–1134, 2003.

[29]. S. Fischer et K.U. Simmer, "Beamforming microphone arrays for speech acquisition in noisy environments," Speech Communication, vol. 20, no. 3–4, pp. 215–227, 1996.

[30]. J. Bitzer, K. U. Simmer et K.-D. Kammeyer., "Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement," in Proc. 24th IEEE Int. Conf. on ASSP, vol. 5, pp. 2965–2968, 1999.

[31]. S. Fischer et K.-D. Kammeyer, "Broadband beamforming with adaptive postfiltering for speech acquisition in noisy environments," in Proc. 22th IEEE Int. Conf. on ASSP, vol. 1, pp. 359–362, 1997.

[32]. C. Marro, Y. Mathieux et K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," IEEE Trans. on SAP, vol. 6, no. 3, pp. 240–259, 1998.

[33]. I. Cohen, "Multichannel post-filtering in nonstationary noise environments," IEEE Trans. on SP, vol. 52, no. 5, pp. 1149–1160, 2004.

[34]. I. Cohen et al., "An integrated real-time beamforming and postfiltering system for nonstationary noise environments," in EURASIP Journal on Applied Signal Processing, pp. 1064–1073, 2003.

[35]. X. Zhang et Y. Jia, "A soft decision based noise cross power spectral density estimation for two-microphone speech enhancement systems," IEEE IC on ASSP, vol. 1, pp. I/813–16, 2005.

[36]. R. Martin, "Noise power spectral estimation based on optimal smoothing and minimum statistics," IEEE Trans. on SAP, vol. 9, pp. 504–512, 2001.

[37]. I. Cohen et B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," IEEE Trans. on SAP, vol. 9, no. 1, pp. 12–15, 2002.

[38]. I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," IEEE Trans. on SAP, vol. 11, no. 5, pp. 466–475, 2003.

[39]. L. V.-Dominguez, "New insights into the high-order Yule-Walker equations," IEEE Trans. on ASSP, vol. 38, pp. 1649–1651, 1990.

[40]. K. C. Sharman et E. Breakenbridge, "Estimation of signal parameters using the maximum likelihood method," IEE Colloq. on Mathematical Aspects of DSP, pp. 8/1–8/4, 1994.

[41]. A. Nehorai et P. Stoica, "Adaptive algorithms for constrained ARMA signals in the presence of noise," IEEE Trans. on ASSP, vol. 36, pp. 1282–1291, 1988.

[42]. S. M. Kay, "The effects of noise on the autoregressive spectral estimator," IEEE Trans. on ASSP, vol. 27, no. 5, pp. 478–485, 1979.

[43]. D. Izraelevitz et J. S. Lim, "Properties of the overdetermined normal equation method for spectral estimation when applied to sinusoids in noise," IEEE Trans. on ASSP, vol. 33, no. 2, pp. 406–412, 1985.

[44]. S. Prasad et K. V. S. Hari, "Improved ARMA spectral estimation using the canonical variate method," IEEE Trans. on ASSP, vol. 35, no. 6, pp. 900–903, 1987.

[45]. H. T. Hu, "Linear prediction analysis of speech signals in the presence of white Gaussian noise with unknown variance," IEE Proc. Vision, Image and Signal Processing, vol. 145, pp. 303–308, 1998.

[46]. Q. Zhao et al., "Improvement of LPC analysis of speech by noise compensation," Electronics and Communications in Japan, vol. 83, pp. 73–83, 2000.

[47]. A. Trabelsi, F.R. Boyer et Y. Savaria, "On the application of minimum noise tracking to cancel cosine shaped residual noise," accessible en ligne à l'adresse URL: http://www.polymtl.ca/biblio/epmrt/rapports/rt2006-09.pdf.

[48]. A. Trabelsi, F.R. Boyer et Y. Savaria, "Speech enhancement based noise PSD estimator to remove cosine shaped residual noise," 50th Midwest Symposium on Circuits and Systems, pp. 393–396, 5–8 Aug. 2007.

[49]. A. Trabelsi, F.R. Boyer et Y. Savaria, "A two-microphone algorithm for speech enhancement," à paraître dans IEEE Transactions on ASLP.

[50]. A. Trabelsi, F.R. Boyer, Y. Savaria et M. Boukadoum, "Improving LPC analysis of speech in additive noise," IEEE Northeast Workshop on Circuits and Systems, pp. 93–96, 5–8 Aug. 2007.

[51]. A. Trabelsi, F.R. Boyer, Y. Savaria et M. Boukadoum, "Iterative noise-compensated method to improve LPC based speech analysis," IEEE Int. Conf. on Electronics, Circuits and Systems, pp. 1364–1367, 11–14 Dec. 2007.

[52]. M.R. Sambur et N.S. Jayant, "LPC Analysis/Synthesis from Speech Inputs Containing Quantizing Noise or Additive White Noise," IEEE Trans. on ASSP, Vol. 24, No. 6, pp. 488–494, 1976.

[53]. V. R. Algazi et al., "Robust LPC analysis and synthesis using the KL transformation of acoustic subwords spectra", IEEE Int. Conf. on ASSP, vol. 1, pp. 468–471, 1989.

[54]. P.C. Checkley et V. Kühnel, "Advantages of an adaptive multi-microphone system," The Hearing Review, vol. 7, no. 5, pp. 58–60 & 74, 2000.

[55]. K. K. Paliwal, "Estimation of noise variance from the noisy AR signal and its application in speech enhancement," IEEE Trans. on ASSP, vol. 36, pp. 292–294, 1988.

[56]. G. H. Golub et C. F. Van Loan, Matrix Computations. Johns Hopkins University Press, New York, 1983.

ANNEXE A

IMPROVING LPC ANALYSIS OF SPEECH IN ADDITIVE NOISE

A. Trabelsi[+1], F.R. Boyer[+1], Y. Savaria[*1] and M. Boukadoum[+2]

Departments of computer[+] and electrical[*] engineering

École Polytechnique de Montréal[1]

Université du Québec à Montréal[2]

Publication source:

IEEE Northeast Workshop on Circuits and Systems,

pp. 93–96, 5–8 Aug. 2007

Improving LPC Analysis of Speech in Additive Noise

A. Trabelsi[✦], F.R. Boyer[✦], and Y. Savaria[✱]
Departments of computer[✦] and electrical[✱] engineering
École Polytechnique de Montréal
Montreal, Quebec, Canada
Abdelaziz.Trabelsi@polymtl.ca

M. Boukadoum
Department of computer science
Université du Québec à Montréal
Montreal, Quebec, Canada
Boukadoum.Mounir@uqam.ca

***Abstract*—Linear prediction based speech (LPC) analysis is known to be sensitive to the presence of additive noise. In this paper, we present a noise-compensated method for LPC analysis which ensures good spectral matching between the original speech spectrum and the autoregressive (AR) model spectrum. In this method, the noise periodogram is obtained first by applying a simplified noise power spectral density (PSD) estimator on the calculated noisy periodogram. Then, the effect of noise on the spectral parameters is decreased by gradually subtracting values of the resulting noise autocorrelation coefficients from the coefficients derived from the noisy speech. By taking the absolute value of the estimated reflection coefficients as the decision criterion, we show that this iterative procedure ensures a significant decrease of the degrading effect of noise while the estimated autocorrelation matrix is guaranteed to be positive definite. The method was tested on real speech signals and yielded superior performance when compared to conventional LPC analysis, even in severe noisy conditions.**

I. INTRODUCTION

LPC is the most common parametric modeling technique for low-bit-rate speech coding and it is a powerful tool in speech analysis. LPC analysis models the speech signal as a p-th order AR system. In a controlled, noise-free environment, the performance of LPC is often satisfactory. With additive noise, however, the signal spectrum is no longer an AR model spectrum [1], and the LPC analysis process yields poor spectral estimates of the input speech. As such, the spectrum of the LPC synthesis filter becomes distorted, and this results in an overall degradation in the quality of the recovered speech.

A wide variety of approaches that aim at improving the noise immunity of LPC analysis have been proposed. A noise reduction approach using the pitch synchronous addition technique was reported in [2]. In that method, the estimated frame is obtained by carrying out a synchronous average of multiple pitch periods included within the analysis frame. Although pitch synchronous analysis yields reliable pitch estimation in a voiced section of speech, it is more vulnerable to errors at voiced-unvoiced boundaries where pitch periods are often irregular. Moreover, the improvement in signal-to-noise ratio (SNR) achieved by the method is constrained by the number of pitch periods included in the analysis frame (of 20-25 ms duration), which is assumed to be stationary.

Autocorrelation subtraction methods based on the common spectral subtraction technique have also been proposed [3]. The approach reduces degradation caused by additive noise by subtracting the noise periodogram from the periodogram of noisy speech after estimation of the former during non-speech activity periods. Whereas this method is efficient when the noise is locally stationary, it becomes ineffective when the noise statistics are time-varying or when the noise power is equal to or greater than the signal power. On the other hand, noise-compensated AR coefficient estimation has been successfully applied to noise reduction in LPC analysis [4]. Noise compensation is achieved by gradually subtracting a noise power estimate from the autocorrelation function (ACF) of noisy speech. In the study, the noise was assumed to be known, and the noise power was reduced at a given iteration step from the whole ACF of the corrupted speech.

The present study proposes an alternative noise-compensated method for LPC analysis which

ensures good spectral matching between the original speech spectrum and the AR model spectrum. In this method, the noise periodogram is obtained first by applying a simplified noise PSD estimator on the calculated noisy periodogram. Then, the effect of noise on the spectral parameters is decreased by gradually subtracting values of the resulting noise autocorrelation coefficients from the coefficients derived from the noisy speech. By taking the absolute value of the estimated reflection coefficients as the decision criterion, we will show that this iterative procedure ensures a significant decrease in the degrading effect of noise while the estimated autocorrelation matrix is guaranteed to be positive definite. Unlike the methods mentioned above, our method properly tracks the noise power level related to the noise autocorrelation coefficients. Thus, the need to consider an estimation error of the noise power is avoided. In addition, the noise is assumed to be totally unknown, which is the case in many speech processing applications.

II. AR MODELING OF SPEECH IN NOISE

Assume that the signal sequence $\{s(n),\ n = 1,2,\ldots,N\}$ is produced at the output of a p-th order AR process driven by a white noise process $w(n)$ with distribution $N\left(0,\sigma_w^2\right)$. We have

$$s(n) = -\sum_{k=1}^{p} a_k s(n-k) + w(n) \tag{1}$$

where $\{a_k,\ k = 1,2,\ldots,p\}$ are real coefficients of the AR process. In most applications, the signal to be modeled contains white noise. Thus, the AR signal $s(n)$ is corrupted by a sequence of white noise $v(n)$ with distribution $N\left(0,\sigma_v^2\right)$ as follows:

$$x(n) = s(n) + v(n) \tag{2}$$

Moreover, the corrupting noise $v(n)$ is assumed to be uncorrelated with the driving noise $w(n)$, i.e., $E\{v(n)w(n-m)\} = 0$ for all m, where $E\{\cdot\}$ is the expected value operator. The order p of the AR process is assumed to be known.

For the noiseless case, $\{a_k\}$ can be found by solving the Yule-Walker equations

$$R_{ss}(k) = -\sum_{m=1}^{p} a_m R_{ss}(k-m),\quad k \geq 1 \tag{3}$$

where $R_{ss}(k)$ denotes the ACF of $s(n)$ and can be estimated using the biased ACF estimator:

$$R_{ss}(k) = \frac{1}{N}\sum_{n=0}^{N-1-k} s(n)s(n+k),$$
$$k = 0,1,\ldots,N-1 \tag{4}$$

For a p-th order AR model, $R_{ss}(k)$ needs to be determined only for $0 \leq k \leq p$. Obviously, any p equations are sufficient to determine the AR parameters. Usually, $k = 0,1,\ldots,p$ is chosen so that a set of symmetric Toeplitz equations is obtained.

When noise is present, although the only accessible observation is $x(n)$ instead of $s(n)$, we can still estimate $R_{ss}(k)$ by noting that

$$R_{ss}(k) = \begin{cases} R_{xx}(0) - \sigma_v^2 & \text{for } k = 0 \\ R_{xx}(k) & \text{for } k \neq 0 \end{cases} \tag{5}$$

Since the noise variance σ_v^2 is assumed to be unknown, $R_{ss}(k)$ can be estimated for all lags other than zero from $x(n)$. This can be accomplished by using the high-order Yule-Walker equations, where $R_{ss}(0)$ is not involved [5]. Unfortunately, such approach suffers from the positive definiteness constraint of the estimated autocorrelation matrix and from the effect of the noise, whose energy spreads all over the autocorrelation lags of speech, i.e., nonstationary noises. The possible singularity of the autocorrelation matrix may lead to a substantial increase in the variance of the AR spectral estimate. Ignoring the noise effect on all lags other than zero may cause underestimation of the noise variance.

In this paper, we consider estimation of the AR parameters in the presence of additive noise. We assume that the noise variance is unknown and that the noise effect extends over the whole autocorrelation function of speech. Note that in practice, linear prediction is equivalent to AR spectral estimation.

III. PROPOSED METHOD

In the previous section, we discussed the problem of AR spectral estimation when white noise is added. In most practical environments, however, speech is degraded by additive noise that is not white. In order to deal with various noise processes, an estimate of the noise variance should be subtracted from the whole ACF $R_{xx}(k)$, k = 0,1,...,p, of speech. Generalizing the result given in (5), we expect the noiseless R_{ss} estimate to be expressed in the form

$$R_{ss}(k) = R_{xx}(k) - \sigma_v^2(k) \cdot u(k),\ k = 0,1,...,p \qquad (6)$$

where $u(.)$ is the discrete-time unit step. Let $R_{nn}(k) = \sigma_v^2(k) \cdot u(k)$, k = 0,1,...,p be the biased ACF estimate of the unknown noise process. Let $P_{nn}(\omega_l)$ be an estimation of noise periodogram at frequency ω_l which can be expressed as

$$P_{nn}(\omega_l) = \sum_{k=-(N-1)}^{N-1} R_{nn}(k) e^{-j\omega_l k} \qquad (7)$$

Because of the efficiency of the minimum statistics algorithm to perform in both stationary and nonstationary noise [6], a simplified noise PSD estimator is used to estimate the noise periodogram P_{nn}. To carry out the running spectral minima search, the D subsequent noise PSD estimates are divided into 2 sliding data subwindows of $D/2$ samples, and the minimum estimate is updated every time instant. Using that running update rate, the highest delay that could occur in response to a rising noise power is about D. Taking the inverse DFT of P_{nn} yields an estimate of the noise ACF.

Consider the nonsingularity constraint of the noiseless autocorrelation matrix derived from (6). In order for R_{ss} to be a positive definite autocorrelation matrix, it is necessary that the associated reflection coefficients be strictly bounded by one in magnitude [7]. Taking this condition as a decision criterion, the effective amount of noise reduction can readily be monitored by means of an iterative processing scheme. In the considered iterative procedure, the update equation for R_{ss} at the ith iteration is as follows:

$$\widetilde{R}_{ss}(k) = \left|R_{xx}(k)\right| - (1 - \mu \cdot i) \cdot \left|R_{nn}(k)\right| \cdot u(k) \Big| \cdot \mathrm{sgn}\left\{\left|R_{xx}(k)\right| - (1 - \mu \cdot i) \cdot \left|R_{nn}(k)\right| \cdot u(k)\right\} \qquad (8)$$

where μ is a step size parameter that must be a positive number. The step size affects both the rate of convergence and the estimation accuracy. This parameter can be optimized by experiments and was set to 0.05 in this work. Notice that the sign function, "sgn", is used in (8) to prevent overestimation of the noiseless $\widetilde{R}_{ss}$, particularly at high- order autocorrelation lags where the noise ACF estimate frequently decays to values below zero.

The steps of the iterative procedure may be summarized as follows:

1. Compute the biased ACF estimate R_{xx} and the corresponding periodogram P_{xx}.
2. Apply the simplified noise PSD estimator to estimate noise periodogram P_{nn}.
3. Compute the estimate of the noise ACF R_{nn} by inverse DFT of P_{nn}.
4. Set the initial iteration value i to 0.
5. Compute the autocorrelation lag values $\widetilde{R}_{ss}(k)$, k = 0,1,...,p using the update equation in (8).
6. Evaluate the prediction and reflections coefficients by the Levinson-Durbin recursion.
7. Let Γ_{j+1} be the $(j+1)$st reflection coefficient, j = 0,1,...,p-1. If $|\Gamma_{j+1}| < 1$ holds for all j, then the prediction coefficients obtained in step 6 are valid and the iterative procedure is complete. Otherwise, increase the iteration value ($i = i+1$) and go back to step 5.

A block diagram of the proposed method is depicted in Figure 1.

IV. PERFORMANCE EVALUATION

To evaluate the effectiveness of the proposed method, we conducted an LPC analysis of three vowels taken from an utterance of the sentence "Flowers grow in the garden" spoken by a male talker. The analysis frame length was set to 256 data samples (23 ms at 11.025 kHz sampling rate) with 50% overlap. The analysis and synthesis windows had thus the perfect reconstruction property. The LPC analysis was performed in the presence of two noise processes, namely white and

impulsive noise, at diverse SNR levels. The input SNR was varied from -5 dB to 15 dB in 5 dB steps. The LPC order was set to 15.

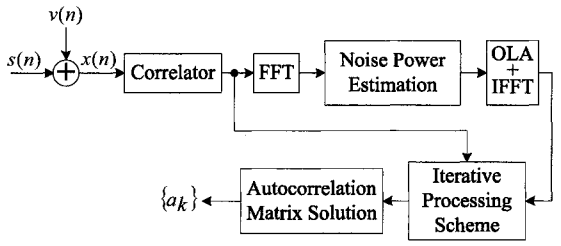

Figure 1. Block diagram of the proposed method.

We evaluated the improvement achieved by our method in terms of cepstral distance according to the following procedure:

- Compute the cepstral distance C_{ci} between the spectrum of the clean speech and that of the improved speech.
- Compute the cepstral distance C_{cd} between the spectrum of the clean speech and that of the degraded speech.
- Compute the difference C_I between C_{ci} and C_{cd}.

We select C_I as the performance evaluation criterion. An improvement is achieved if $C_I < 0$, and the spectral estimation becomes poorer if $C_I > 0$. Tables I and II summarize the results of the cepstral distance in the estimation of the noiseless ACF when white noise and impulsive noise are added, respectively. Notice that in the calculation of the cepstral distance, the initial cepstral coefficient, which represents the average energy of the speech frame, was discarded. From the obtained results, it is seen that the proposed method is effective at decreasing the variance of the estimated prediction coefficients, even in severe noisy conditions. It can also be noted that the degree of improvement in the case of white noise is higher than that of impulsive noise. This is because the noise PSD estimator performs better in the presence of stationary noise as opposed to nonstationary noise.

TABLE I. PERFORMANCE EVALUATION IN TERMS OF CEPSTRAL DISTANCE (WHITE NOISE)

SNR (dB)	-5	0	5	10	15
Vowel /o/	-1.32	-1.87	-2.15	-1.93	-1.75
Vowel /i/	-0.69	-0.81	-1.05	-1.37	-1.53
Vowel /a/	-1.42	-2.05	-2.31	-2.69	-2.94

TABLE II. PERFORMANCE EVALUATION IN TERMS OF CEPSTRAL DISTANCE (IMPULSIVE NOISE)

SNR (dB)	-5	0	5	10	15
Vowel /o/	-0.53	-0.84	-1.08	-0.95	-0.79
Vowel /i/	-0.31	-0.43	-0.58	-0.64	-0.73
Vowel /a/	-0.64	-1.02	-1.34	-1.69	-1.88

The robustness of the proposed procedure to additive white noise is illustrated in Figure 2. The figure shows the LPC power spectra estimates obtained by averaging 10 realizations for the second /o/ vowel of the sentence under examination before and after noise compensation. The input SNR was set to 5 dB. It can be seen from curves (b) and (c) that the latter tracks the LPC structure of the clean signal more closely than the former. In particular, the shape of the first three formants is better preserved in the improved section (c) compared to the degraded section (b). The behavior of the first 3 formants is perceptually of crucial importance in many applications, e.g., voice coders [8].

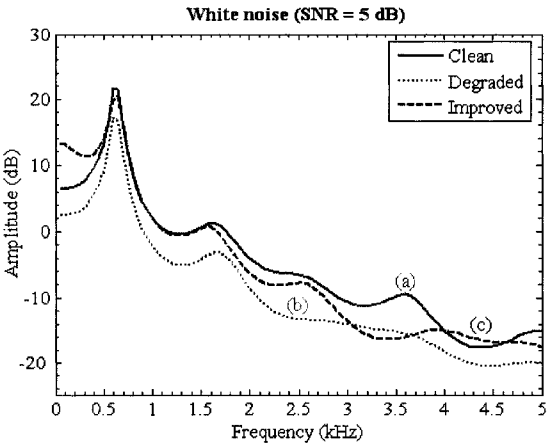

Figure 2. Averaged LPC spectra of (a) clean signal (solid), (b) degraded signal (dotted), and (c) improved signal (dashed), in the presence of white noise (Input SNR = 5 dB).

V. DISCUSSION

We will now address the issue of the computational effort needed by the proposed method. Clearly, the major computational effort in the proposed method is the iteration of the Levinson-Durbin recursion. Note that only the Levinson-Durbin recursion is iterated and the computation of the autocorrelation lag values is not considered. For a p-th order model, the Levinson-Durbin recursion requires $O(p^2)$ arithmetic operations. When the step size parameter μ is set to 0.05, only 20 iterations for all autocorrelation lags are needed in the worst case. Assuming $(p + 1)$ autocorrelation lags and setting $p = 15$, the iterative procedure requires about p^3 arithmetic operations in comparison to about p^2 operations for standard LPC spectral analysis. For a signal of length N, computing the autocorrelation function requires about $N \cdot (p + 1)$ arithmetic operations. Therefore, if $N >> p$, the cost associated with finding the autocorrelation function will dominate the computational effort of the modeling procedure.

VI. SUMMARY

We have developed an iterative method to compensate for the degrading effect of noise on the prediction parameters when an LPC spectral analysis is used. It was verified that an accurate estimate of noise variance prior to LPC analysis maximizes the improvement in the AR spectral estimate. It was also observed that subtracting an estimate of the noise power from the whole autocorrelation function of speech, allows dealing with more general noise processes than white noise. In the future, it is planned to perform a comparative study of our method with other existing noise compensation algorithms.

REFERENCES

[1] S. M. Kay, "The effect of noise on the autoregressive spectral estimator," *IEEE Trans. on ASSP*, vol. 27,. pp. 478-485, 1979.

[2] Y. Kuroiwa et al., "An improvement of LPC based on noise reduction using pitch synchronous addition," *ISCAS'99*, vol. 3, pp. 122-125, 1999.

[3] C. Un et al., "Improving LPC analysis of noisy speech by autocorrelation subtraction method," *IEEE Int. Conf. on ASSP*, vol. 6, pp. 1082-1085, 1981.

[4] Z. Qifang et al., "Improvement of LPC analysis of speech by noise compensation," *Electron Comm Jpn*, P.3, vol. 83, pp. 73-83, 2000.

[5] S. Kay, "Noise Compensation for Autoregressive Spectral Estimates," *IEEE Trans. on ASSP*, vol. 28,. pp. 292-303, 1980.

[6] R. Martin, "Noise power spectral estimation based on optimal smoothing and minimum statistics," *IEEE Trans. on SAP*, vol. 9,. pp. 504-512, 2001.

[7] M.H. Hayes, *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, Inc, 1996.

[8] D. O'Shaughnessy, *Speech Communications: Human and Machine*, 2nd ed. New York: IEEE Press, 2000.

ANNEXE B

ITERATIVE NOISE-COMPENSATED METHOD TO IMPROVE LPC BASED SPEECH ANALYSIS

A. Trabelsi[+1], F.R. Boyer[+1], Y. Savaria[*1] and M. Boukadoum[+2]

Departments of computer[+] and electrical[*] engineering

École Polytechnique de Montréal[1]

Université du Québec à Montréal[2]

Publication source:

IEEE Int. Conf. on Electronics, Circuits and Systems,

pp. 1364–1367, 11–14 Dec. 2007

Iterative Noise-Compensated Method to Improve LPC Based Speech Analysis

A. Trabelsi[1], F.R. Boyer[1], and Y. Savaria[2]
Departments of computer[1] and electrical[2] engineering
École Polytechnique de Montréal
Montreal, Quebec, Canada
Abdelaziz.Trabelsi@polymtl.ca

M. Boukadoum
Department of computer science
Université du Québec à Montréal
Montreal, Quebec, Canada
Boukadoum.Mounir@uqam.ca

***Abstract*—It is well known that linear predictive coding (LPC) performs well when the prediction coefficients are estimated from noise-free speech, and the system tends to degrade and perform poorly on noisy speech. This paper describes a method to minimize the degradation on the prediction coefficients in the presence of noise when an LPC analysis is used. In this method, a more accurate estimation of noise power is computed by using a simplified noise power spectral density (PSD) estimator. After an inverse discrete Fourier transform (DFT), the extracted noise autocorrelation coefficients are gradually subtracted from the coefficients derived from noisy speech according to an iterative processing scheme. The proposed processing scheme takes the absolute value of the estimated reflection coefficients as the decision criterion. It is shown that performing this iterative procedure on every autocorrelation lag ensures a substantial decrease in the degrading effects of noise, while the estimated autocorrelation matrix is guaranteed to be positive-definite. Experimental results indicate that the variance of the estimated prediction coefficients can be decreased significantly using the proposed method.**

I. INTRODUCTION

Accurate estimation of LPC coefficients (or spectral parameters) is an important problem in low-bit-rate speech coding. Because the coefficients are directly related to the pole locations, which are function principally of formant frequencies and bandwidths, the standard LPC technique requires that the spectral parameters be estimated from noise-free speech. In noisy conditions, however, the LPC coefficients are subject to severe temporal variations as compared to those observed when processing noise-free speech. Thus, they may no longer represent the proper configurations and shapes of the glottal source and the vocal tract system. Consequently, the spectrum of the LPC synthesis filter exhibits formant shifting and bandwidth widening, leading to an overall degradation in the quality of the reconstructed speech.

There have been several methods that aim to reduce the effects of noise on LPC parameters. The existing methods used to retrieve the spectral parameters of speech from noise corrupted measurements can be divided into two main categories: autoregressive moving average (ARMA) process based estimation and parameter compensation [1]. The basic principle of the ARMA process based estimation is to represent the p-th order noisy autoregressive (AR) model of speech by an ARMA (p, p) process and then to estimate the AR parameters by using the selected ARMA process. Attributed to this category are the modified Yule-Walker (MYW) equations method [1] and the recursive prediction error (RPE) method [2]. Although the MYW method yields a straightforward algorithm from the computational point of view, it suffers from poor estimation accuracy and relatively low efficiency due to the use of high order autocorrelation lag estimates. On the contrary, the RPE method yields consistent parameter estimates at the cost of high computational complexity.

On the other hand, iterative [3] and adaptive [4] noise subtraction methods have been suggested to compensate the spectral parameters for the noise bias. These methods can be attributed to the

parameter compensation category. Noise compensation in [3] is achieved by gradually subtracting a noise power estimate from the autocorrelation function (ACF) of noisy speech. In this study, the noise was assumed to be known, and the noise power was reduced at a given iteration step from the whole ACF of the corrupted speech. Instead of deriving the exact noise variance, the method in [4] determines a suitable bias that should be subtracted from the zero-lag autocorrelation function. In this method, the LPC synthesis filter is guaranteed to be stable by confining the noise variance to be less than the minimum eigenvalue of the autocorrelation matrix.

This paper describes an alternative method to minimize the degradation on the prediction coefficients in the presence of noise when an LPC based speech analysis is used. In this method, a more accurate estimation of noise power is computed by using a simplified noise PSD estimator. After an inverse DFT, the extracted noise autocorrelation coefficients are gradually subtracted from the coefficients derived from noisy speech according to an iterative processing scheme. The proposed processing scheme takes the absolute value of the estimated reflection coefficients as the decision criterion. In contrast to the parameter compensation methods mentioned above, the proposed iterative procedure is performed on every autocorrelation lag and allows a substantial decrease in the degrading effects of noise on the spectral parameters at the expense of an increase in the computational effort. In addition, the noise is assumed to be totally unknown, which is the case in many speech processing applications. Unlike the method we initially proposed in [5], the amount of noise power subtracted from each autocorrelation lag may be different. As a result, better noise compensation can be achieved, while the estimated autocorrelation matrix is always guaranteed to be positive-definite.

II. NOISE COMPENSATION IN THE CONTEXT OF LINEAR PREDICTION

Let $s(n)$ be the discrete-time series of a speech signal to be estimated from a noise corrupted measurement

$$x(n) = s(n) + v(n), \quad n = 1,\ldots,N \tag{1}$$

where $v(n)$ is uncorrelated white noise process with unknown variance σ_v^2.

If $s(n)$ satisfies the "all-pole" assumption, it can be modeled as the output of a p-th order linear predictor process excited by a sequence of zero-mean white noise $w(n)$ with variance σ_w^2:

$$s(n) = -\sum_{k=1}^{p} a_k s(n-k) + w(n) \tag{2}$$

In the noise-free case, the p unknown parameters $\{a_k\}$ of the linear predictor can be obtained by solving the Yule-Walker equations

$$R_{ss}(k) + \sum_{m=1}^{p} a_m R_{ss}(k-m) = 0, \quad k = 1,2,\ldots,p \tag{3}$$

where $R_{ss}(k)$ denotes the ACF of the process $s(n)$, and can be estimated using the biased ACF estimator

$$\hat{R}_{ss}(k) = \frac{1}{N}\sum_{n=0}^{N-1-k} s(n)s(n+k), \; k = 0,1,\ldots,N-1 \tag{4}$$

For a p-th linear predictor, $R_{ss}(k)$ needs to be determined only for $0 \le k \le p$. If the process $s(n)$ is replaced by $x(n)$, then the ACF of the latter, $\hat{R}_{xx}(k) = E\{x(n+k)x(n)\}$ may be expressed as

$$\hat{R}_{xx}(k) = \hat{R}_{ss}(k) + \hat{R}_{vv}(k) = \hat{R}_{ss}(k) + \sigma_v^2\delta(k) \tag{5}$$

where $\hat{R}_{vv}$ is the ACF of the white noise process $v(n)$, $\delta(.)$ is the Kronecker delta, and $E\{\cdot\}$ is the expected value operator. Given that only $x(n)$ is available and the noise variance is unknown, $\hat{R}_{ss}(k)$ cannot be evaluated for $0 \le k \le p$ directly from (5). One way around this difficulty is to use the high-order Yule-Walker equations, where $\hat{R}_{ss}(0)$ is not involved, for the evaluation of $\hat{R}_{ss}(k)$ from $x(n)$ for all lags other than zero [6]. Unfortunately, such approach suffers severely from the positive definiteness constraint of the estimated autocorrelation matrix and from the effect of the noise, whose energy spreads all over the autocorrelation lags of speech, i.e., nonstationary

noises. The possible singularity of the autocorrelation matrix may lead to a substantial increase in the variance of the estimated spectral parameters. Ignoring the noise effect on all lags other than zero may cause underestimation of the noise power.

It is the purpose of this paper to deal with the evaluation of the noiseless $\hat{R}_{ss}$ when the noise power is unknown and when the noise effects extend over the whole autocorrelation function of speech.

III. Proposed Method

Although the discussion in the previous section is only concerned with the problem of estimating the prediction parameters for a process $s(n)$ in white noise $v(n)$, by subtracting an estimate of the noise power from $\hat{R}_{xx}(k)$, $k = 0,1,\ldots,p$ we may easily extend these results to more general noise processes than white noise. Generalizing the result given in (5), we expect the noiseless $\hat{R}_{ss}$ estimate to be expressed in the form

$$\hat{R}_{ss}(k) = \hat{R}_{xx}(k) - \sigma_v^2(k)u(k),\ k = 0,1,\ldots,p \qquad (6)$$

where $u(.)$ is the discrete-time unit step. Let $\hat{R}_{nn}(k) = \sigma_v^2(k)u(k)$, $k = 0,1,\ldots,p$ be the biased ACF estimate of the unknown noise process. Let $\hat{P}_{nn}$ be an estimation of the unknown noise power over the frequency range of interest.

In spite of the efficiency of the minimum statistics algorithm to perform in both stationary and nonstationary noise [7], a simplified noise PSD estimator is used to estimate the noise power, $\hat{P}_{nn}$. To carry out the running spectral minima search, the D subsequent noise PSD estimates are divided into 2 sliding data subwindows of $D/2$ samples, and the minimum estimate is updated every time instant. Using that running update rate, the highest delay that could occur in response to a rising noise power is about D. Taking the inverse DFT of $\hat{P}_{nn}$ and using the biased ACF estimator yields a more accurate estimate of the noise ACF.

Consider the nonsingularity constraint of the noiseless autocorrelation matrix derived from (6). In order for $\hat{R}_{ss}$ to form a positive-definite autocorrelation matrix it is necessary that the associated reflection coefficients be strictly bounded by one in magnitude [8]. Taking this condition as a decision criterion, the effective amount of noise reduction can readily be monitored by means of an iterative processing scheme. In the considered iterative procedure, the update equation for $\hat{R}_{ss}$ at the ith iteration is as follows

$$\hat{R}_{ss}(k) = \left|\tilde{R}_{ss}(k)\right| \cdot \mathrm{sgn}\left\{\tilde{R}_{ss}(k)\right\} \qquad (7a)$$

where

$$\tilde{R}_{ss}(k) = \left|\hat{R}_{xx}(k)\right| - (1 - \mu \cdot i) \cdot \left|\hat{R}_{nn}(k)\right| \cdot u(k) \qquad (7b)$$

and where μ is the step size parameter to adjust a trade-off between the convergence speed and the estimation accuracy. This parameter can be optimized by experiments and was set to 0.05 in this work. Note that the sign function, "sgn", in (7) is used to prevent an overestimation of the noiseless $\hat{R}_{ss}$, in particular at high order autocorrelation lags where the noise ACF estimate frequently decays to values below zero. It is also considered to perform the iterative procedure on every autocorrelation lag, starting from lag zero. As a result, different amount of noise power may be subtracted from each autocorrelation lag, which can considerably decrease the degrading effects of noise on the spectral parameters at the expense of an increase in computational effort. The steps of the iterative procedure may be summarized as follows:

1. Compute the sliding window FFT analysis of the noise corrupted measurement.
2. Apply the simplified noise PSD estimator to evaluate the noise power, $\hat{P}_{nn}$.
3. Compute the estimate of the noise ACF, $\hat{R}_{nn}$ by inverse DFT of $\hat{P}_{nn}$ and by using (4).
4. Compute the biased ACF estimate, $\hat{R}_{xx}$.
5. Set the initial autocorrelation lag value k to 0.
6. Set the initial iteration value i to 0.
7. Compute the autocorrelation lag value $\hat{R}_{ss}(k)$, using (7).
8. Set $\hat{R}_{ss}(m) = \vec{R}_{act}(m)$, for $0 \le m \ne k \le p$.

9. Evaluate the prediction and reflection coefficients by the Levinson-Durbin recursion.
10. Let Γ_{j+1} be the $(j+1)$st reflection coefficient, $j = 0,1,\ldots,p-1$. If $|\Gamma_{j+1}| < 1$ holds for all j, then the prediction coefficients obtained in step 9 are valid, set $\vec{R}_{act}(k) = \hat{R}_{ss}(k)$, increase the autocorrelation lag value ($k = k+1$) and go to step 11. Otherwise, increase the iteration value ($i = i+1$) and go back to step 7.
11. If $k \leq p$, then go back to step 6 and repeat the process. Otherwise, the noiseless $\hat{R}_{ss}$ estimate is found and the iterative procedure is complete.

Note that in the above description of the iterative procedure, we make reference to the running ACF estimate, $\vec{R}_{act}$. At the startup of the procedure, such ACF is initialized to $\vec{R}_{act}(k) = \hat{R}_{xx}(k)$, $k = 0,1,\ldots,p$.

IV. Experimental Results

The performance of the proposed method was investigated using real speech signals filtered at 11.025 kHz sampling rate. The analysis frame length was set to 256 data samples (23 ms) with 50% overlap. The analysis and synthesis windows had the perfect reconstruction property. The experimental results were obtained by processing an utterance of the sentence "Flowers grow in the garden" spoken by a male talker and corrupted with two noise processes, namely white and impulsive, at diverse SNR levels. The input SNR was varied from -5 dB to 15 dB in 5 dB steps. The LPC spectral analysis was performed using a 15-pole predictor.

Figure 1 shows the LPC power spectra estimates obtained by averaging 10 realizations for the second /o/ vowel of the sentence under examination before and after noise compensation. The input SNR was set to 0 dB (white noise). It can be seen from curves (b) and (c) that the latter tracks the LPC structure of the clean signal more closely than the former. In particular, the shape of the first three formants is better preserved in the improved section (c) compared to the degraded section (b). The behavior of the first 3 formants is perceptually of crucial importance in many applications, e.g., formants vocoders. Similar results were obtained in Figure 2 when impulsive noise is added to the speech signal.

Let the cepstral distance between the spectrum of the clean speech and that of the improved speech be denoted by C_{ci}. Similarly, let the cepstral distance between the spectrum of the clean speech and that of the degraded speech be denoted by C_{cd}. We select $C_I = C_{ci} - C_{cd}$ as the performance evaluation in terms of cepstral distance of the proposed iterative procedure. An improvement is achieved if $C_I < 0$, and the spectral estimation becomes poorer if $C_I > 0$. Table I summarizes the results of the cepstral distance in the estimation of the noiseless ACF when white noise is added. Notice that in the calculation of the cepstral distance, the initial cepstral coefficient, which represents the average energy of the speech frame, was discarded. From these results, it can be concluded that the proposed method is effective in decreasing the variance of the estimated prediction parameters even in severe noisy conditions.

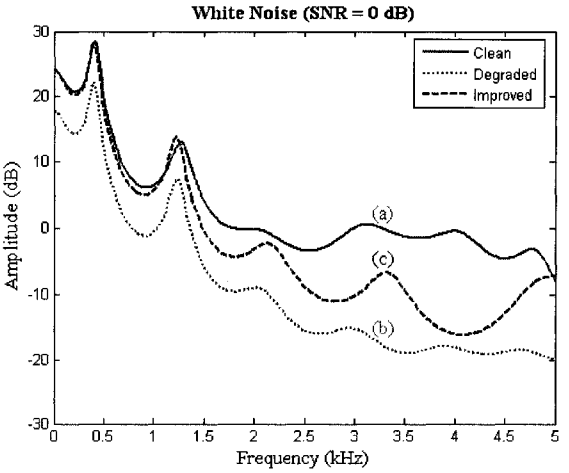

Figure 1. LPC spectra of (a) clean signal (solid), (b) degraded signal (dotted), and (c) improved signal (dashed), in the presence of white noise (Input SNR = 0 dB).

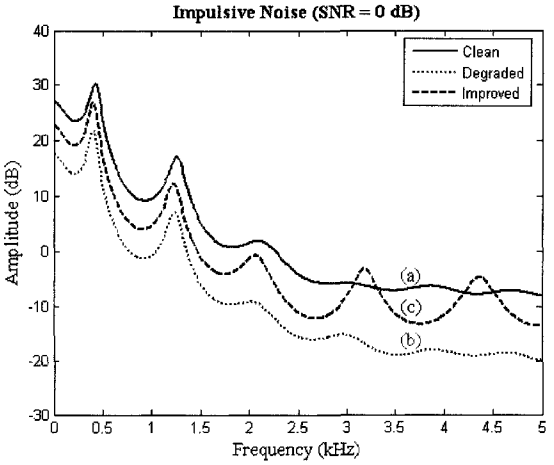

Figure 2. LPC spectra of (a) clean signal (solid), (b) degraded signal (dotted), and (c) improved signal (dashed), in the presence of impulsive noise (Input SNR = 0 dB).

TABLE I. PERFORMANCE EVALUATION IN TERMS OF CEPSTRAL DISTANCE (WHITE NOISE)

SNR (dB)	-5	0	5	10	15
/o/	-1.37	-1.91	-2.19	-1.97	-1.81
/i/	-0.73	-0.85	-1.09	-1.41	-1.55
/a/	-1.45	-2.08	-2.35	-2.72	-2.98

V. DISCUSSION

We now look at the computational complexity of the proposed method. Obviously, the major computational effort is the iteration of the Levinson-Durbin recursion. An approach to quantify this computational effort is to estimate the number of arithmetic operations (multiplications and divisions) required to perform the method within a given analysis frame (ignoring all additions and subtractions). Note that only the Levinson-Durbin recursion is iterated and the computation of the autocorrelation lag values is not considered. For a p-th order model, the Levinson-Durbin recursion requires $O(p^2)$ arithmetic operations. When the step size parameter μ is set to 0.05, only 20 iterations per autocorrelation lag are needed in the worst case. Assuming $(p + 1)$ autocorrelation lags and setting $p = 15$, the iterative procedure requires about p^4 arithmetic operations in comparison to about p^2 operations for standard LPC spectral analysis. While significant, the extra computational effort required by the proposed method should not be a problem for speech signals given their relatively low bandwidth and the increasing use of parallel processing architectures in signal processing applications.

VI. SUMMARY

In summary, we conclude that the proposed method is effective at estimating the noiseless ACF of a noisy speech when the noise power is unknown and when the degrading effects of noise extend all over the autocorrelation function. However, the reduction in the noise power is achieved at the cost of an increase in computational effort. In the future, we will quantify the computational burden more precisely by performing a comparative study of how our method performs with respect to other existing noise compensation algorithms.

REFERENCES

[1] S. M. Kay, *Modern spectrum estimation*. Englewood Cliffs, NJ. Prentice-Hall, 1988.

[2] A. Nehorai and P. Stoica,. "Adaptive algorithms for constrained ARMA signals in the presence of noise," *IEEE Trans. on ASAP*, vol. 36,. pp. 1282-1291, 1988.

[3] Z. Qifang et al., "Improvement of LPC analysis of speech by noise compensation," *Electron Comm Jpn*, P.3, vol. 83, pp. 73-83, 2000.

[4] H. T. Hu, "Linear prediction analysis of speech signals in the presence of white Gaussian noise with unknown variance," *IEE Proc.-Vis. Image Signal Process.*, vol. 145,. pp. 303-308, 1998.

[5] A. Trabelsi, F.R. Boyer, M. Boukadoum, and Y. Savaria, "Improving LPC Analysis of Speech in Additive Noise," *IEEE Int. Midwest Symp. on Circuits and Systems*, 2007.

[6] S. Kay, "Noise Compensation for Autoregressive Spectral Estimates," *IEEE Trans. on ASSP*, vol. 28,. pp. 292-303, 1980.

[7] R. Martin, "Noise power spectral estimation based on optimal smoothing and minimum statistics," *IEEE Trans. on SAP*, vol. 9,. pp. 504-512, 2001.

[8] M.H. Hayes, *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, Inc, 1996.